



**A University of Sussex PhD thesis**

Available online via Sussex Research Online:

<http://sro.sussex.ac.uk/>

This thesis is protected by copyright which belongs to the author.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Please visit Sussex Research Online for more information and further details

# Design and Characterisation of Tool Inhibitors of DNA Damage Response Proteins

Katie Melissa Duffell

DPhil (Doctor of Philosophy)

University of Sussex

September 2017



## I Declaration

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Signed .....

Katie Melissa Duffell

All work presented in this thesis was conducted by me except for the cases outlined below:

- Small molecules tested in this work were synthesised at the Sussex Drug Discovery Centre, or the Automated High-Throughput Medicinal Chemistry and Synthesis Lab, University of Sussex.
- NMR fragment screening was conducted at the University of Birmingham.
- X-ray diffraction data was collected and analysed by Dr Mark Roe, X-Ray Crystallography Collaborative Research Facility, University of Sussex.
- His6-tagged SMC5 head domain and His6-tagged KAT2A BRD constructs were devised by Dr Aaron Alt and Dr Antony Oliver respectively, Genome Damage and Stability Centre, University of Sussex.
- KNIME workflow used to identify compounds available from preferred suppliers was devised by Dr Ben Wahab, Sussex Drug Discovery Centre, University of Sussex.

## II Acknowledgements

Firstly, I would like to thank Prof. Simon Ward, for allowing me the opportunity to tackle this enthralling project and providing the scientific guidance and unfading support and encouragement I needed to do so. You have been a truly inspirational supervisor.

I would like to thank the Medical Research Council for funding my studies.

Thanks go to our collaborators at the SGC who provided valuable expert advice, data and small molecules.

I would like to thank the staff at the University of Sussex. Special thanks go to my thesis committee, Prof. Antony Carr, Prof. Jessica Downs, Dr Frances Pearl and Dr Jo Murray, for the stimulating discussions and helpful advice; to Dr Aaron Alt, Dr Mohan Rajasekaran and Dr Antony Oliver for assistance with protein production and assay advice, to Dr Chrisostomos Prodromou for guidance with ITC, and to Dr Mark Roe for performing X-ray crystallography experiments. It has been a pleasure to work with you all. Also, thank you to Alethea, Alex and Paul for their help.

A big thank you to my colleagues from the SDDC, past and present, for their help and encouragement. Particular thanks to Prof. John Atack, Dr Sarah Walker and Dr Darren Le Grand for their kind guidance and support, to Dr Ben Wahab for computational advice, to Dr Lewis Pennicott and all the chemists who synthesised compounds for this work, and to the Oncology team who have provided much helpful scientific discussion and cake.

I would also like to thank my lovely friends and lab buddies; Hayley, Oran, Tom, Rhiannon, Jess, Irene, Adam, Dan, Dave, Poj, Gavin, Mark, Sean, James, Tori, Lucas, Chloe, Thalia, James, Fiona, Rachael, Yusuf, Ryan, Tristan, Iain and Gareth, who have been a tremendous help to me in making it through these past four years. I will miss you all dearly. Special thanks go to Tori, Tom and Gareth for their help proofreading.

Finally, a huge thank you to my incredible family; my husband Ash, my parents, Jeff and Annette, and my sister Claire, without whose support, love and encouragement I could not have achieved this.

For Nan.



### III Abstract

As a result of the high cost of drug discovery, it is imperative that promising targets with strong disease association are identified and validated before embarking on costly molecule discovery and development phases. Chemical inhibitors provide an excellent tool for target validation. This project was initiated with the aim of designing and characterising tool inhibitors of proteins involved in the DNA damage response, to enable detailed mechanistic biological investigation, disease validation and initiation of translational drug discovery projects.

The SMC complexes are critically important in coordinating chromosome condensation, sister chromatid cohesion, DNA repair, homologous recombination and transcriptional regulation. They represent interesting oncology targets and are compelling targets for tool inhibitor development. The core SMC proteins and NSE subunits of SMC5/6 and cohesin SMC1 and SMC3 were modelled and their druggabilities assessed, with an aim to advance to inhibitor development for any of the proteins that proved druggable. Unfortunately, the SMC5/6 and cohesin complexes were established as not druggable.

BAF180 represents a major clear cell renal cell carcinoma (ccRCC) cancer gene, exhibiting truncating mutations in 41% of samples in a series of primary ccRCCs.<sup>1</sup> It has therefore been highlighted as a promising opportunity to target ccRCC using a synthetic lethal therapeutic approach.<sup>2</sup> Hopkins *et al.* utilised a novel screening technique to identify several genes that are synthetic lethal with BAF180.<sup>3</sup> In this work, the druggabilities of three of these genes were assessed, and KAT2A was identified as a promising target for tool inhibitor development.

Druggable pockets were identified at the KAT2A histone acetyltransferase (HAT) domain and bromodomain. The catalytically active HAT domain was prioritised as the preferred target. Potential KAT2A HAT domain inhibitors were available from a high-throughput screen (HTS), which utilised a fluorescence based activity assay.<sup>4,5</sup> Unfortunately, this assay format proved prohibitively unreliable, and moreover, after discounting the hits from the HTS as likely false positives, the inventory of KAT2A inhibitors was quickly exhausted and the target abandoned.

Focus turned to the KAT2A bromodomain. Available X-ray crystal structures were utilised in a computational drug design effort, and T<sub>m</sub> shift, TR-FRET, ITC and X-ray crystallography techniques were optimised and established in-house to enable characterisation of prospective KAT2A tool inhibitors. Over 3000 small molecules and fragments were screened, and an assortment of novel KAT2A bromodomain binders were identified. The optimised assays, novel chemical matter and ligand-bound crystal structures afford an exciting opportunity to develop potent and selective KAT2A bromodomain tool inhibitors.

## IV Contents

I	Declaration.....	1
II	Acknowledgements.....	2
III	Abstract.....	3
IV	Contents.....	4
V	Abbreviations.....	8
1	Introduction .....	16
1.1	<b>The DNA Damage Response</b> .....	17
1.1.1	<i>DNA Damage Detection</i> .....	17
1.1.2	<i>DNA Damage Signalling</i> .....	20
1.1.3	<i>DNA Damage Repair</i> .....	21
1.2	<b>DNA Repair Dysregulation in Cancer</b> .....	25
1.2.1	<i>Hallmarks of Cancer</i> .....	25
1.2.2	<i>Genome Instability</i> .....	31
1.3	<b>Targeted Cancer Therapy</b> .....	32
1.3.1	<i>Synthetic Lethality</i> .....	33
1.3.2	<i>Chemical Synthetic Lethal Screens</i> .....	36
1.3.3	<i>Genetic Synthetic Lethal Screens</i> .....	38
1.4	<b>Target Identification</b> .....	44
1.4.1	<i>Exploratory Research</i> .....	44
1.4.2	<i>Target Viability Data</i> .....	45
1.4.3	<i>Structure-Based Druggability Assessment</i> .....	46
1.4.4	<i>Modelling Three-Dimensional Protein Structure</i> .....	49
1.5	<b>Target Validation</b> .....	52
1.5.1.	<i>Genetic Modulation</i> .....	53
1.5.2.	<i>mRNA Modulation</i> .....	55
1.5.3.	<i>Protein Modulation</i> .....	57
1.6	<b>Project Aim</b> .....	60
2	Materials and Methods.....	61
2.1	<b>Bioinformatics</b> .....	61
2.1.1	<i>Data Procurement</i> .....	61
2.1.2	<i>Homology Searches</i> .....	61
2.1.3	<i>Sequence Alignments</i> .....	61
2.1.4	<i>Conservation Scoring</i> .....	61
2.2	<b>Homology Modelling</b> .....	62

2.2.1	<i>Model Building</i>	62
2.2.2	<i>Model Preparation and Minimisation</i>	64
<b>2.3</b>	<b>Structural Alignment and Visualisation</b>	64
<b>2.4</b>	<b>Pocket Prediction and Scoring</b>	65
<b>2.5</b>	<b>Data Handling - KNIME</b>	65
2.5.1	<i>Filter Results of HTS and Hit Confirmation Assay</i>	65
2.5.2	<i>Retrieve and React Isothiazolones</i>	67
2.5.3	<i>Generate Conformers</i>	68
2.5.4	<i>Select Highest Scoring Docking Pose and Retrieve Top 40% of Compounds</i>	68
2.5.5	<i>Generate Final List of Docked Hit Compounds</i>	69
<b>2.6</b>	<b>Data Analysis and Visualisation</b>	70
<b>2.7</b>	<b>Molecular Docking</b>	70
2.7.1	<i>KAT2A HAT Domain</i>	70
2.7.2	<i>KAT2B Bromodomain</i>	71
<b>2.8</b>	<b>Protein Techniques</b>	71
2.8.1	<i>Transformation and Expression</i>	71
2.8.2	<i>Protein Purification</i>	72
<b>2.9</b>	<b>Screening Compounds</b>	73
<b>2.10</b>	<b>Isothermal Titration Calorimetry</b>	74
2.10.1	<i>SMC5 Head Domain Protein</i>	74
2.10.2	<i>KAT2A BRD Protein</i>	74
<b>2.11</b>	<b>HAT Activity Assays</b>	75
2.11.1	<i>Fluorescence Based Activity Assay</i>	75
2.11.2	<i>EpiQuik™ Activity Assay</i>	75
<b>2.12</b>	<b>KAT2A BRD Tm Shift Assay</b>	76
<b>2.13</b>	<b>X-Ray Crystallography</b>	76
2.13.1	<i>Crystal Growing in Hanging Drops</i>	76
2.13.2	<i>Crystal Soaking</i>	77
2.13.3	<i>Co-Crystallisation</i>	77
2.13.4	<i>Diffraction Data Collection, Structure Solution and Refinement</i>	77
<b>2.14</b>	<b>KAT2A BRD TR-FRET Assay</b>	79
2.14.1	<i>Assay Design</i>	80
2.14.2	<i>Assay Optimisation and Screening</i>	80
<b>3</b>	<b>Assessing the SMC Complexes as Potential Targets for Tool Inhibitor Development</b>	81
<b>3.1</b>	<b>Introduction</b>	81
3.1.1	<i>Cohesin</i>	86

3.1.2	<i>Condensin</i> .....	89
3.1.3	<i>SMC5/6</i> .....	91
<b>3.2</b>	<b>Results</b> .....	96
3.2.1	<i>SMC5 and SMC6</i> .....	96
3.2.2	<i>Non-SMC Elements</i> .....	105
3.2.3	<i>Cohesin</i> .....	108
<b>3.3</b>	<b>Discussion</b> .....	110
4	Assessing the Druggability of Three Targets that are Synthetic Lethal with BAF180 .....	113
<b>4.1</b>	<b>Introduction</b> .....	113
4.1.1	<i>Chromatin Remodelling Complexes</i> .....	113
4.1.2	<i>BAF180</i> .....	115
4.1.3	<i>RNF4</i> .....	117
4.1.4	<i>ASF1A</i> .....	118
4.1.5	<i>GCN5/KAT2A</i> .....	119
<b>4.2</b>	<b>Results</b> .....	122
4.2.1	<i>RNF4</i> .....	123
4.2.2	<i>ASF1A</i> .....	124
4.2.3	<i>GCN5/KAT2A</i> .....	125
<b>4.3</b>	<b>Discussion</b> .....	128
4.3.1	<i>KAT2A-BAF180 Synthetic Lethality</i> .....	129
5	Towards Development of KAT2A HAT Domain Tool Inhibitors .....	132
<b>5.1</b>	<b>Introduction</b> .....	132
5.1.1	<i>Phylogeny, Structure and Function</i> .....	132
5.1.2	<i>Existing Inhibitors</i> .....	137
5.1.3	<i>New Chemical Matter</i> .....	141
<b>5.2</b>	<b>Results</b> .....	141
5.2.1	<i>Selectivity Potential</i> .....	141
5.2.2	<i>Sifting the Existing Chemical Matter</i> .....	142
5.2.3	<i>Fluorescence Based Activity Assay</i> .....	143
5.2.4	<i>EpiQuik™ HAT Activity Assay</i> .....	150
5.2.5	<i>Isothiazolones as Potential HAT Inhibitors</i> .....	151
<b>5.3</b>	<b>Discussion</b> .....	152
6	Design and Characterisation of KAT2A Bromodomain Tool Inhibitors .....	155
<b>6.1</b>	<b>Introduction</b> .....	155
6.1.1	<i>Function and Phylogeny</i> .....	156
6.1.2	<i>Structure and Interactions</i> .....	160

6.1.3	<i>Existing Inhibitors</i> .....	162
6.1.4	<i>Recent KAT2A/B Chemical Matter</i> .....	167
<b>6.2</b>	<b>Results</b> .....	168
6.2.1	<i>Selectivity Potential</i> .....	168
6.2.2	<i>Screening Cascade</i> .....	168
6.2.3	<i>Structure-Based Design From SL1126</i> .....	169
6.2.4	<i>T<sub>m</sub> Shift Optimisation and Screening</i> .....	174
6.2.5	<i>NMR Preparation and Screening</i> .....	182
6.2.6	<i>ITC Hit Confirmation</i> .....	183
6.2.7	<i>ITC Screening</i> .....	187
6.2.8	<i>X-Ray Crystallography</i> .....	194
6.2.9	<i>TR-FRET Assay Design</i> .....	198
6.2.10	<i>TR-FRET Optimisation and Screening</i> .....	206
<b>6.3</b>	<b>Discussion</b> .....	217
<b>7</b>	<b>Discussion</b> .....	220
<b>7.1</b>	<b>SMC Complexes</b> .....	220
<b>7.2</b>	<b>BAF180 Synthetic Lethal Targets</b> .....	221
<b>7.3</b>	<b>KAT2A HAT Domain</b> .....	221
<b>7.4</b>	<b>KAT2A Bromodomain</b> .....	222
<b>VI</b>	<b>References</b> .....	224
<b>VII</b>	<b>Appendix</b> .....	255
<b>VII.I.</b>	<b>SMC Multiple Sequence Alignments</b> .....	255
<b>VII.II.</b>	<b>Histone Acetyltransferase Domain Multiple Sequence Alignment</b> .....	264
<b>VII.III.</b>	<b>Bromodomain Multiple Sequence Alignments</b> .....	265
<b>VII.IV.</b>	<b>Quinolinone Screening Libraries</b> .....	269
<b>VII.V.</b>	<b>NMR Fragment Screening</b> .....	278

## V Abbreviations

AAK1	Adaptor-associated kinase 1
ABC	ATP-binding cassette
ABL	Abelson murine leukaemia viral oncogene homologue
ADME	Absorption, distribution, metabolism and excretion
AGO2	Argonaute-2
AID	Assay Identification
ALT	Alternative lengthening of telomeres
AMP	Adenosine monophosphate
AMP-PNP	Adenosine 5'-( $\beta,\gamma$ -imido)triphosphate
aNHEJ	Alternative NHEJ
APB	ALT-associated PML body
APE1	Apurinic-apyrimidinic endonuclease 1
APL	Acute promyelocytic leukaemia
AP-site	Abasic site
ARID	AT rich interactive domain
ASF1A	Anti-silencing function protein 1 homologue A
ASH1L	Absent, small and homeotic disks protein 1 homologue
ASO	Antisense oligonucleotide
ATAC	ADA2A-containing
ATAT1	Alpha-tubulin N-acetyltransferase 1
ATM	Ataxia-telangiectasia mutated
ATP	Adenosine 5'-triphosphate
ATR	ATM- and Rad3-related
ATRIP	ATR interacting protein
BAC	Bacterial artificial chromosome
BAF	BRG1- or BRM-associated factor
BAF180	BRG1-associated factor 180
BAH	Bromo-adjacent homology
Bak	Bcl-2 homologous antagonist/killer
Bax	Bcl-2 Associated X
BAZ	Bromodomain adjacent to zinc finger domain
BCL	B-cell lymphoma
Bcl-2	B-cell lymphoma-2
BCR	Breakpoint cluster region
BER	Base excision repair
BET	Bromodomain and extra-terminal domain
BH3	Bcl-2 homology-3
BLAST	Basic Local Alignment Search Tool
BLM	Bloom syndrome protein
BRCA1	Breast cancer type 1 susceptibility protein
BRCA2	Breast cancer type 2 susceptibility protein
BRD	Bromodomain
BRG1	Brahma Related Gene 1
BRIT1	BRCT-repeat inhibitor of hTERT expression

BRM	Brahma
BRPF1	Bromodomain And PHD Finger Containing 1
CAF-1	Chromatin assembly factor 1
CAK	Cdk-activating-kinase
CAP	Chromosome-associated proteins
Cas9	CRISPR-associated protein 9
CBP	CREB Binding Protein
CCG	Chemical Computing Group
ccRCC	Clear cell Renal Cell Carcinoma
Cdc7	Cell division control protein 7
Cdk	Cyclin-dependent kinase
CDK1	Cyclin-dependent kinase 1
cDNA	Complementary DNA
CECR2	Cat Eye syndrome chromosome region, candidate 2
CENPI	Centromere protein I
CFA	Complete Freund's Adjuvant
CHARMM	Chemistry at HARvard Macromolecular Mechanics
CHK1	Checkpoint kinase 1
CHK2	Checkpoint kinase 2
cNHEJ	Classical NHEJ
CoA	Coenzyme A
COMA	Ctf19-Okp1-Mcm21-Ame1
CREB	cAMP response element-binding protein
CREBBP	CREB Binding Protein
CRISPR	Clustered regularly interspaced short palindromic repeat
CtBP	C-terminal-binding protein
CTCF	CCCTC-binding factor
CtIP	CtBP-interacting protein
Dbf4	Dumbbell forming protein 4
DDR	DNA-damage response
DDT	DNA damage tolerance
DELTA-BLAST	Domain Enhanced Lookup Time Accelerated BLAST
DGCR8	DiGeorge syndrome critical region 8
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNA2	DNA replication ATP-dependent helicase/nuclease
DNA-PK	DNA-dependent protein kinase
DNA-PKcs	DNA-dependent protein kinase catalytic subunit
dNTP	Deoxynucleotide triphosphate
DOPE	Discrete Optimised Protein Energy
dRP	Deoxyribose phosphate
dRPaase	Deoxyribophosphodiesterase
DSB	Double-stranded break
dsRNA	Double-stranded RNA
DTT	Dithiothreitol
DUB	Deubiquitination

E value	Expectation value
E. coli	Escherichia coli
EBI	European Bioinformatics Institute
ECO1	Establishment of cohesion protein 1
ELISA	Enzyme-linked immunosorbent assay
EMBL	European Molecular Biology Laboratory
EMT	Epithelial-mesenchymal transition
ERCC1	Excision repair cross-complementing rodent repair deficiency, complementation group 1
ES	Embryonic stem
ESCO	Establishment of cohesion 1
ET	Extra-terminal
EXO1	Exonuclease 1
FA	Fanconi anaemia
FAAP24	Fanconi anaemia core complex-associated protein 24
FAN1	Fanconi-associated nuclease 1
FANC	Fanconi anaemia protein
FAT	FRAP-ATM-TRRAP
FATC	FAT carboxy-terminal
FDA	Food and Drug Administration
FEN1	Flap endonuclease 1
FRAP	FKBP12-rapamycin complex-associated protein
FRET	Fluorescence Resonance Energy Transfer
GCN5	General Control Non-depressible protein 5
GEMM	Genetically engineered mouse model
GNAT	GCN5-related N-acetyltransferases
GOI	Gene of interest
GPCR	G-protein-coupled receptor
GST	Glutathione S-transferase
GTP	Guanosine 5'-triphosphate
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HEAT	Huntingtin, Elongation factor 3, protein phosphatase 2A, TOR1
HER3	Human Epidermal Growth Factor Receptor 3
HIF2 $\alpha$	Hypoxia-inducible factor 2 alpha
HIRA	Histone cell cycle regulator A
HIV-1	Human immunodeficiency virus type 1
HMG	High-mobility group
HP1	Heterochromatin protein 1
HPLC	High-performance liquid chromatography
HR	Homologous recombination
HTRF	Homogeneous time-resolved fluorescence
HTS	High-Throughput Screening
IC <sub>50</sub>	Half maximal inhibitory concentration
ICL	Interstrand crosslink
IFN- $\beta$	Interferon- $\beta$



IMAC	Immobilised metal affinity chromatography
iPSC	Induced pluripotent stem cell
IPTG	Isopropyl- $\beta$ -D-thiogalactopyranoside
IR	Infrared
ITC	Isothermal titration calorimetry
KAT2A	Lysine Acetyltransferase 2-A
KAT2B	Lysine Acetyltransferase 2-B
K <sub>d</sub>	Equilibrium dissociation constant
KIF4A	Kinesin Family Member 4A
KNIME	Konstanz Information Miner
KOMP	Knockout Mouse Project
LC-MS	Liquid chromatography-mass spectrometry
LE	Ligand efficiency
LIG	DNA ligase
LOH	Loss of heterozygosity
loxP	Locus of X-over P1
mAbs	Monoclonal antibodies
MAGE	Melanoma-associated antigen
MAPK	Mitogen-activated protein kinase
MAP <sub>POD</sub>	Maximal affinity predicted for a passively absorbed oral drug
MB	$\alpha$ -methylene- $\gamma$ -butyrolactone
MCPH1	Microcephalin 1
MD	Molecular dynamics
MDC1	Mediator of DNA damage checkpoint protein 1
Merlin	Moesin-ezrin-radixin-like protein
MHF	FANCM-associated histone fold protein
miRNA	MicroRNA
MLH	MutL homologue
MLL	Mixed-lineage leukaemia
MMR	Mismatch repair
MMS21	Methyl methanesulfonate-sensitivity protein 21
MOE	Molecular Operating Environment
Mre11	Meiotic recombination 11
MRN	Mre11-Rad50-Nbs1
mRNA	Messenger RNA
MSH	MutS homologue
MutL	Mutator L
MutS	Mutator S
MWCO	Molecular weight cut-off
MYST	MOZ, YBF2/SAS3, SAS2 and TIP60
NAR	Nucleic Acids Research
Nbs1	Nijmegen breakage syndrome protein 1
NCBI	National Centre for Biotechnology Information
NCGC	NIH Chemical Genomics Center
NEBD	Nuclear envelope breakdown
NER	Nucleotide excision repair

NF2	Neurofibromin-2
NF- $\kappa$ B	Nuclear factor- $\kappa$ B
NHA	Non-hydrogen atoms
NHEJ	Nonhomologous end-joining
NIH	National Institutes of Health
NK	Natural killer
NMP	N-Methyl-2-pyrrolidone
NMR	Nuclear magnetic resonance
NSE	Non-SMC element
nt	Nucleotide
NuRD	Nucleosome remodelling and deacetylase
NUT	Nuclear protein in testis
OCT4	Octamer-binding protein 4
PAI1	Plasminogen activator inhibitor 1
PAINS	Pan-assay interference compounds
PARP1	Poly [ADP-ribose] polymerase 1
PBAF	Polybromo-associated BAF
PBRM1	Protein polybromo-1
PCAF	P300/CREBBP-associated factor
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEA3	Polyomavirus enhancer activator 3
PEG	Polyethylene Glycol
PHD	Plant homeodomain
PHYRE	Protein Homology/analogy Recognition Engine
PI3K	Phosphatidylinositol 3-kinase
PIKK	Phosphatidylinositol-3 kinase-related kinase
PLC $\gamma$ 1	Phospholipase C $\gamma$ 1
PML	Promyelocytic leukaemia
PMS2	Post-meiotic segregation protein 2
Pol $\delta$	DNA polymerase $\delta$
Pol $\epsilon$	DNA polymerase $\epsilon$
Polv	DNA polymerase $\nu$
PPI	Protein-protein interaction
PROTAC	Proteolysis Targeting Chimera
PSSM	Position Specific Scoring Matrix
PTEN	Phosphatase and tensin homologue
PTM	Post-translational modification
R&D	Research and Development
Ran	Ras-related Nuclear protein
RAR $\alpha$	Retinoic acid receptor alpha
RB	Retinoblastoma-associated
Rb	Retinoblastoma
rDNA	Ribosomal DNA
RFB	Replication fork barrier

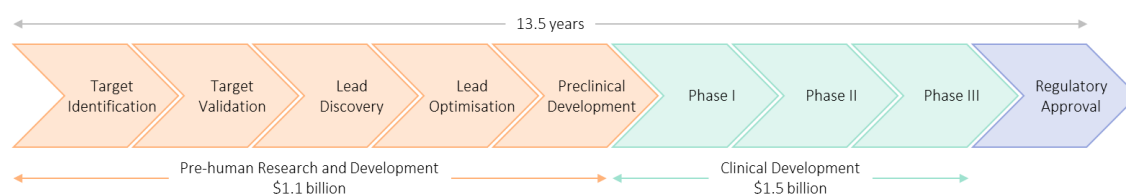
RFC	Replication factor C
RING	Really Interesting New Gene
RISC	RNA-induced silencing complex
RMCE	Recombinase-mediated cassette exchange
RNA	Ribonucleic acid
RNAi	RNA interference
RNase	Ribonuclease H
RNF168	RING finger protein 168
RNF4	RING finger protein 4
RNF8	RING finger protein 8
rNMP	Ribonucleoside monophosphate
Ro5	Rule of 5
RPA	Replication protein A
RPA32	Replication protein A 32 kDa subunit
RPD3	Reduced potassium dependency 3
RSC	Remodelling the structure of chromatin
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
<i>S. pombe</i>	<i>Schizosaccharomyces pombe</i>
SA	Streptavidin
SA1	Stromal Antigen 1
SA2	Stromal Antigen 2
SAGA	Spt-Ada-Gcn5 acetyltransferase
SAHA	Suberoylanilide Hydroxamic Acid
SAR	Structure-activity relationship
SCC	Sister chromatid cohesion protein
SCI	Sister chromatid intertwining
ScpB	Segregation and condensation protein B
SD	Standard Deviation
SDF	Structure-data file
SDSA	Synthesis-dependent strand annealing
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SET	Su(var)3-9, Enhancer-of-zeste and Trithorax
SGA	Synthetic genetic array
SGC	Structural Genomics Consortium
shRNAs	Short hairpin RNAs
SILAC	Stable isotope labelling by amino acids in cell culture
siRNA	Small Interfering RNA
SMC	Structural maintenance of chromosomes
SNURF	Small nuclear RING finger protein
SPR	Surface plasmon resonance
SPT	Suppressor of Ty
SS18	Synovial sarcoma translocated to chromosome 18
SSA	Single-strand annealing
SSB	Single-stranded break
ssDNA	Single-stranded DNA

STAGA	SPT3-TAF9-GCN5 acetyltransferase
STD	Saturation Transfer Difference
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
SUMO	Small ubiquitin-like modifier
SWI/SNF	SWItch/Sucrose Non-Fermentable
TAF	TBP-associated factor
TALEN	Transcription activator-like effector nuclease
TBP	TATA-box binding protein
TCEP	Tris(2-carboxyethyl)phosphine
TEV	Tobacco etch virus
TFIID	Transcription factor II D
TFIIH	Transcription factor II H
TFIIIC	Transcription factor for polymerase III C
TFTC	TBP-free-TAF-containing
TGF- $\beta$	Transforming growth factor- $\beta$
TIP60	Tat-interactive protein, 60 kDa
T <sub>m</sub>	Melting temperature
TNFR	Tumour necrosis factor receptor
Top2	Topoisomerase II
TopBP1	Topoisomerase 2-binding protein 1
TOR1	Target of rapamycin kinase 1
TR-FRET	Time-Resolved FRET
TRIM	Tripartite motif-containing
Tris	Tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
TRPS1	Tricho-rhino-phalangeal syndrome type I protein
TRRAP	Transformation/transcription domain-associated protein
TSP1	Thrombospondin-1
UniProt	Universal Protein Resource
UniProtKB	UniProt Knowledgebase
USP22	Ubiquitin-specific protease 22
UV	Ultraviolet
VEGF-A	Vascular Endothelial Growth Factor-A
VEGFR	Vascular Endothelial Growth Factor Receptors
VHL	Von Hippel-Lindau
Wapl	Wings apart-like protein homologue
WaterLOGSY	Water-Ligand Observed via Gradient Spectroscopy
WPF	Tryptophan-Proline-Phenylalanine
XLF	XRCC4-like factor
XP	Xeroderma pigmentosum, complementation protein
XPF	Xeroderma pigmentosum, complementation protein group F
XRCC1	X-ray repair cross-complementing protein 1
XRCC4	X-ray repair cross-complementing protein 4
ZEB	Zinc finger E-box-binding homeobox
ZFN	Zinc-finger nuclease
ZMYND	Zinc finger Myeloid, Nervy, and DEAF-1 domain-containing protein

911	Rad9-Rad1-Hus1
53BP1	p53-binding protein 1

# 1 Introduction

The discovery of new drugs is labour-intensive and expensive. It is estimated that the fully capitalised cost of development of a new approved drug is more than \$2.5 billion<sup>6</sup>, and takes approximately 13.5 years<sup>7</sup>, Figure 1.1. The incredibly high cost is due in large part to poor clinical success rates, with only ~12% of drugs which enter phase I ultimately being approved.<sup>6</sup> The majority of drugs fail in the clinic for one of two reasons; either they do not work, often due to poor target identification and inadequate validation, or they are not safe.<sup>8</sup> In addition, once in the marketplace, many drugs fall short of recovering their development costs and others must ultimately be withdrawn.<sup>9</sup> As a result, despite significant scientific, technological and managerial advances, the productivity of research and development (R&D) is decreasing. Since 1950, the number of new US Food and Drug Administration (FDA)-approved drugs per billion US dollars spent has halved every nine years.<sup>10</sup> This fall in productivity has been specifically attributed to four factors. Firstly, as the catalogue of approved medicines expands there are increasingly high demands on new drugs. This is an extension of the ‘low-hanging fruit’ problem<sup>11</sup>, whereby therapeutics for easy-to-treat disease targets have been identified and this reduces the efficiency of further research on this disease due to raised requirements of any new drugs. This serves to deter R&D in these areas and encourage researchers towards harder-to-treat diseases. Secondly, authorities such as the FDA are progressively wary of drug safety and this lower tolerance of risk understandably increases the cost of R&D. Thirdly, there has been a tendency to add resources and increase spending on R&D, likely driven by precedented good returns on investment and competition to launch first. Finally, there has been a tendency to overtrust basic research and progress quickly to ‘brute force’ high-throughput screening (HTS) methods rather than adopting an iterative approach. Without thorough target validation or proof-of-concept trials many drugs fail in clinic due to poor efficacy, and while HTS is effective in many cases and offers rapid results, it is expensive and often yields hits with high molecular weight, poor ligand efficiency (LE) and limited potential for optimisation.<sup>10,12,13</sup>



**Figure 1.1. Model of Research and Development.** R&D timeline with associated costs to launch one new drug.<sup>6</sup>

Clearly drug safety matters and it would be irresponsible to suggest that regulations should be relaxed for the comparatively trivial purpose of improving R&D productivity. As such, it seems efficiency must be improved by adopting a more disciplined approach to drug discovery, in early

development in particular.<sup>14</sup> Resources and spending should be tightly managed, the ‘high-hanging fruit’ should be targeted by taking advantage of appropriate technological advances, targets should be carefully selected and well validated and the physicochemical properties of candidate compounds should be assessed prior to clinical development in an attempt to anticipate safety issues.<sup>13</sup> Fundamentally, costly drug discovery programmes should only be undertaken when there is good confidence that regulation of the target protein will result in the desired therapeutic effect. Target identification and validation are therefore critically important.

Drug discovery begins with a hypothesis derived from basic research predicting that regulating a particular protein will afford a therapeutic effect for a disease state. This protein is then assessed as a potential drug target. As well as evoking the desired biological response, a good target should be ‘druggable’. This requires that it is accessible to a conceivable drug molecule and binding of such a molecule should induce a measureable response. If the protein is deemed a viable target it must be scrutinised to validate the legitimacy of the therapeutic effect. A range of validation techniques are available and typically a multi-validation approach is favourable. Particularly popular approaches utilise antisense technology, small interfering RNA (siRNA), transgenic animals, monoclonal antibodies or small bioactive molecules as tool inhibitors.<sup>8</sup>

## 1.1 The DNA Damage Response

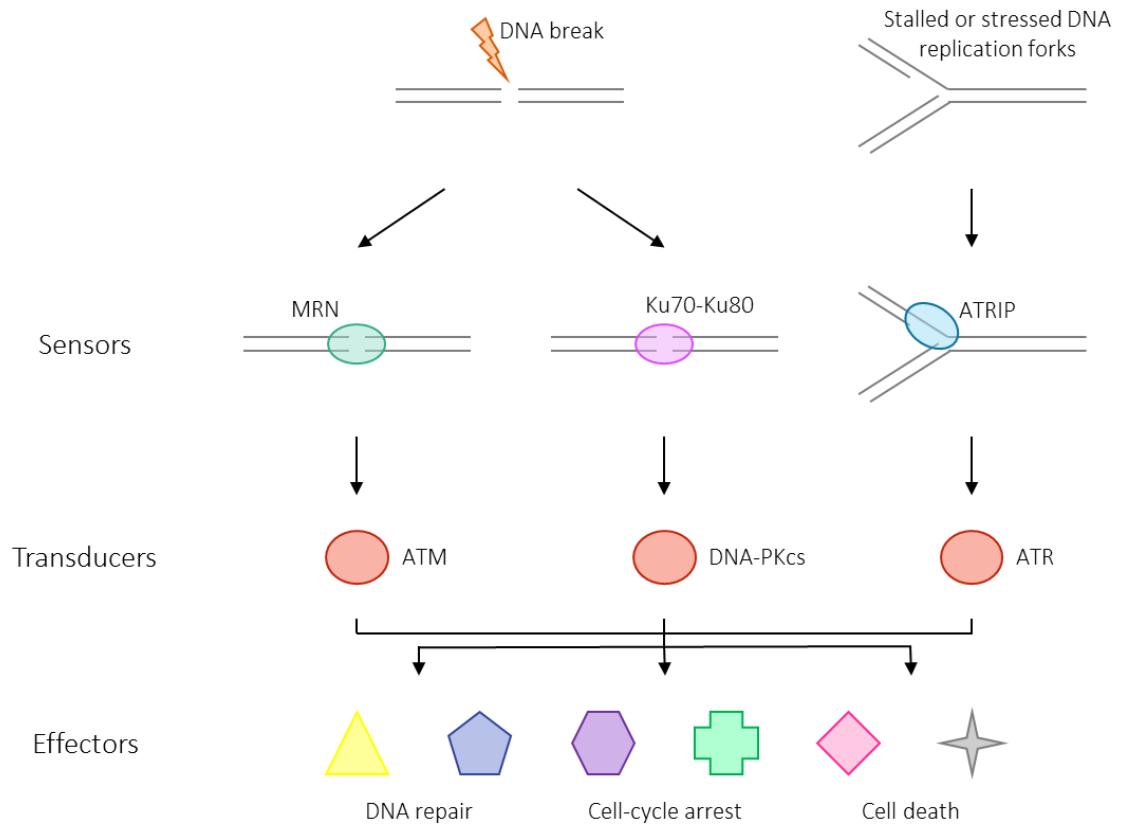
R&D productivity varies between different disease areas. Fortunately, basic oncology research appears to be delivering new and improved targets for which effects observed in the lab correlate better with efficacy in humans. Furthermore, authorities are more tolerant of risk on the understanding that it is outweighed by benefits, and the current catalogue of approved cancer therapeutics leaves considerable scope for improvement.<sup>10</sup> As DNA repair plays such a prominent role in tumourigenesis and tumour response to therapy, the DNA damage response network has been identified as a promising target for improved medicines.

Every cell in the human body acquires tens of thousands of DNA lesions each day<sup>15</sup>, capable, if not correctly repaired, of blocking genome replication and transcription or leading to mutations and aberrations that threaten cell viability.<sup>16</sup> Thankfully cells have developed mechanisms to combat the threats posed by DNA damage. They are able to detect DNA lesions, signal their presence and promote their repair, and arrest cell cycle progression while this is achieved.<sup>17,18</sup> These mechanisms are collectively referred to as the DNA-damage response (DDR).<sup>16</sup>

### 1.1.1 DNA Damage Detection

The DDR kinase signalling pathway comprises of a protein kinase cascade with mediator proteins which facilitate phosphorylation events, Figure 1.2.<sup>18</sup> In mammalian cells the most upstream

DDR kinases are the ATM (ataxia-telangiectasia mutated), ATR (ATM- and Rad3-related) and DNA-PKcs (DNA-dependent protein kinase catalytic subunit). These large serine/threonine kinases are members of the phosphatidylinositol-3 kinase-related kinase (PIKK) family.<sup>19</sup> As is currently understood, DNA-PKcs and ATM are primarily involved in repairing double-stranded breaks (DSBs), whereas ATR can respond to a wider range of DNA lesions.<sup>20</sup>



**Figure 1.2. Framework of the DDR Kinase Signalling Pathway.** The DDR pathway is comprised of sensor proteins that recognise damaged DNA, transducer proteins that relay and amplify the signal and effector proteins that control DNA repair, cell cycle progression and cell death. Figure adapted from Maréchal & Zou, 2013 and Falck *et al.*, 2005.<sup>18,21</sup>

PIKKs all have similar domain organisations; their kinase domains are located near the carboxyl termini, flanked by FAT (FRAP-ATM-TRRAP) and FATC (FAT carboxy-terminal) domains, and the amino terminal and internal regions are composed of numerous HEAT (Huntington-elongation factor 3-protein phosphatase 2A-TOR1) repeats, Figure 1.3.<sup>22</sup> It is suggested that the HEAT repeats serve to interact with DNA<sup>23</sup> and mediate protein-protein interactions important for the function and regulation of the kinases. For example ATM interacts with Nbs1 (Nijmegen breakage syndrome protein 1), a component of the MRN (Mre11-Rad50-Nbs1) complex, via specific internal HEAT repeats<sup>21,24</sup>, and ATR with ATRIP (ATR interacting protein) via amino terminal repeats.<sup>25</sup> The FAT and FATC domains act as regulatory domains. In human cells in the absence of DNA damage the ATM FAT and kinase domains interact, inhibiting kinase activity. When DNA is damaged, ATM Ser1981 is autophosphorylated. This disrupts the interaction and



alleviates inhibition of ATM.<sup>26</sup> Analogously, in response to DNA damage the ATR FAT domain is phosphorylated at Thr1989 to promote the stimulation of ATR by TopBP1 (Topoisomerase 2-binding protein 1).<sup>27</sup> The FATC domains are similarly responsible for kinase activity.<sup>28</sup>



**Figure 1.3. PIKK Domain Organisation.** Generalised schematic of functional domains of PIKKs, annotated to show approximate number of amino acids. Figure adapted from Maréchal & Zou, 2013.<sup>18</sup>

DSBs are the most perilous form of DNA damage as they can enable genome rearrangements, and unrepaired DSBs are typically lethal.<sup>29</sup> When a DSB occurs one of the first factors recruited is the MRN complex. It appears that MRN recognises blunt double-stranded DNA (dsDNA) ends or dsDNA ends with short single-stranded DNA (ssDNA) overhangs, i.e. dsDNA/ssDNA junctions, characteristic of DSBs, and recruits and activates ATM.<sup>30</sup> The activation of ATM results in phosphorylation of as many as 700 substrates such as p53, CHK2, BRCA1 (breast cancer type 1 susceptibility protein) and H2AX, which mediate a multitude of effects, for example on cell cycle arrest, DNA repair, chromatin restructuring, protein synthesis or even apoptosis.<sup>31</sup>

DNA-PKcs is also recruited to DSBs, here by the Ku70-Ku80 heterodimer, to form the active DNA-PK complex.<sup>32</sup> Interestingly the C-terminal region of Ku80, which is thought to be necessary for interaction with DNA-PKcs, is homologous to the conserved region in the C-terminus of Nbs1 that serves to recruit ATM.<sup>21</sup> It seems that DNA-PKcs functions chiefly in nonhomologous end-joining and regulates only a small number of targets, compared to the hundreds of proteins phosphorylated in an ATM- or ATR-dependent manner.<sup>33</sup>

ATR is recruited to sites of DNA damage by association with its binding partner ATRIP. The ATR-ATRIP complex is localised to stalled forks via a direct interaction between ATRIP and replication protein A (RPA)-coated ssDNA<sup>34</sup>, which is generated during DNA replication when there is a loss of coordination between the helicase and DNA polymerase such that the polymerase stalls but the helicase does not.<sup>35</sup> This ATRIP-mediated ATR localisation is necessary but not sufficient for activation of the ATR kinase. In addition, a second complex forms on the RPA-coated ssDNA, independent of ATR and ATRIP, at ssDNA/dsDNA junctions. This complex comprises of the ring-shaped Rad9-Rad1-Hus1 (911) clamp, which interacts with the dsDNA; Rad17-replication factor C (RFC), the clamp loader; and TopBP1, which interacts with the ssDNA.<sup>18,35,36</sup> As the complex is assembled, the TopBP1 ATR activation domain is revealed, which enables interaction with the ATR-ATRIP complex and activates the kinase.<sup>37,38</sup> ATR activation results in phosphorylation of a plethora of substrates including CHK1 and RPA32<sup>39</sup>, which again mediate an array of effects.

### 1.1.2 DNA Damage Signalling

As discussed, the PIKKs regulate many hundreds of downstream repair proteins. The kinases control DNA repair by three dominant mechanisms. Firstly, they directly regulate the activity of DNA repair enzymes via post-translational modifications. Secondly, they initiate modification of the chromatin near a DNA lesion, which creates a local environment permissive for repair and serves as a scaffold to localise additional repair effectors. Finally, the kinases act at a global level to establish a cellular environment conducive to repair. This may include regulating the cell cycle, chromosome mobility or deoxynucleotide triphosphate (dNTP) levels.<sup>20</sup> One example of each of these control mechanisms is discussed below.

ATR kinase has a critical function in initiating interstrand crosslink (ICL) repair in eukaryotic cells. ICLs are highly damaging because they prevent separation of the two DNA strands, which causes DNA replication and transcription to halt. Unfortunately, they are also perhaps the most difficult lesions to repair, requiring specialised repair mechanisms governed by the FANCs (Fanconi anaemia proteins), in addition to input from nucleotide excision repair and DSB repair.<sup>20</sup> Understanding of how DNA ICLs are removed has been greatly enhanced by the study of patients with Fanconi anaemia (FA), a rare genetic disorder that leads to ICL sensitivity.<sup>40</sup> ICLs are initially recognised by FANCM, a DNA translocase, which associates with FAAP24 (Fanconi anaemia core complex-associated protein 24) and two histone-like factors, MHF-1 and -2 (FANCM-associated histone fold proteins 1 and 2), to form the FA anchor complex. FANCM then recruits the FA core complex, a nine-membered E3 ubiquitin ligase. FANCI and FANCD2 are also recruited to the DNA damage site, binding directly to the DNA, and are mono-ubiquitinated by this FA core complex. The mono-ubiquitinated FANCI-FANCD2 complex serves as a scaffold for the recruitment of further ICL repair proteins including nucleases such as FAN1 (Fanconi-associated nuclease 1), XPF (Xeroderma pigmentosum, complementation protein group F) and SLX4, and polymerases such as Polv.<sup>40,41</sup> ATR controls many of the events in the FA pathway. For example, it is responsible for phosphorylation of several FA proteins including FANCM, FANCA and FANCG, found in the FA core complex, and FANCI and FANCD2.<sup>20</sup> The phosphorylation of FANCI appears especially critical for FA pathway activation as it is required for mono-ubiquitination and localisation of FANCD2.<sup>42</sup> In this way, ATR directly regulates the activity of these DNA repair enzymes via post-translational modification.

DDR kinases induce several chromatin changes that promote a local environment conducive to repair. These include creating a platform for recruiting effectors, enabling access to the DNA and inhibiting nearby transcription, which could interfere with repair. Subsequent to its initial activation at DNA ends, ATM triggers a host of events on the chromatin flanking a DSB. Pivotal

to these is phosphorylation of histone variant H2AX on serine 139 in the C-terminal tail to form  $\gamma$ H2AX.<sup>43</sup> MDC1 (mediator of DNA damage checkpoint protein 1) recognises and binds to the phosphorylated H2AX and Nbs1 as well as ATM. The complex of  $\gamma$ H2AX, MDC1 and MRN thereby serves to recruit additional ATM to chromatin flanking the DSB. This results in propagation of the  $\gamma$ H2AX mark to a domain incorporating over 500 kb of chromatin termed  $\gamma$ H2AX foci.<sup>18</sup> The  $\gamma$ H2AX-MDC1 forms a scaffold for recruitment of further signalling and effector proteins. For example, RING (Really Interesting New Gene) ubiquitin ligases RNF8 (RING finger protein 8) and RNF168 (RING finger protein 168) are recruited, which trigger a ubiquitination cascade.<sup>44</sup> Downstream this ubiquitination regulates the recruitment of BRCA1 and 53BP1 (p53-binding protein 1), which are believed to regulate the repair choice between nonhomologous end-joining and homologous recombination.<sup>45</sup> Phosphorylation of H2AX also induces changes to chromatin structure by recruiting chromatin remodelling complexes such as SWI/SNF (SWItch/Sucrose Non-Fermentable). SWI/SNF chromatin remodelling activity is targeted to DSBs by H3 acetylation by GCN5 (General Control Non-depressible protein 5)<sup>46</sup> and by interaction with BRIT1 (BRCT-repeat inhibitor of hTERT expression), which binds  $\gamma$ H2AX after damage.<sup>47</sup> Relaxation of the chromatin near the break by SWI/SNF improves access for DNA repair proteins to the site of damage.<sup>20</sup>

Finally, the PIKKs also facilitate repair at a more global level by establishing a conducive cellular environment. For example, the replication stress checkpoint is mediated by ATR. When DNA damage is detected ATR phosphorylates and activates CHK1 (checkpoint kinase 1), which inhibits cell cycle progression to allow time to repair the damage prior to replication or mitosis.<sup>20,48</sup>

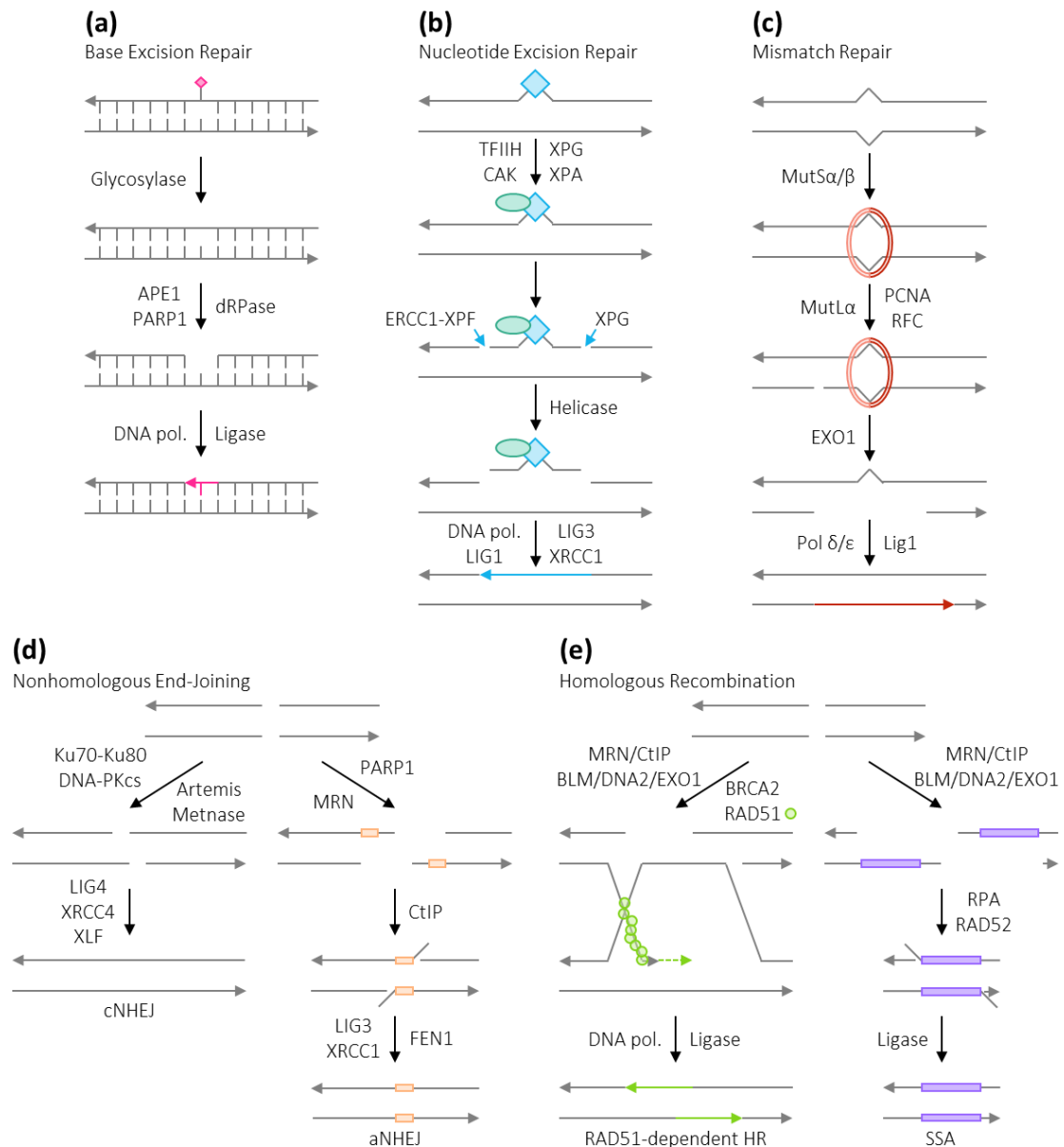
### 1.1.3 DNA Damage Repair

There are five major classes of DNA repair pathways: base excision repair (BER), nucleotide excision repair (NER) and mismatch repair (MMR), which act on damage present on single strands; and nonhomologous end-joining (NHEJ) and homologous recombination (HR), which act at DSBs, Figure 1.4.<sup>29</sup> These pathways can act independently at specific DNA lesions but in reality often function in complex networks. For example, if repair by one pathway is unsuccessful, a lesion may be passed to a second pathway, or if DNA repair by one pathway causes successive damage, this might require repair by another. These repair networks constitute a robust system for maintaining genomic stability.<sup>29</sup>

#### *Base Excision Repair*

BER describes a collection of mechanisms for the repair of base lesions that cause little distortion to the DNA helix structure. DNA damage from oxidation, deamination and alkylation causes

these non-bulky lesions such as ring-opened bases and small adducts such as oxidised bases. The lesions are recognised by one of at least 11 distinct DNA glycosylases and the damaged base is removed, leaving an abasic site (AP-site).<sup>49</sup> Apurinic-apyrimidinic endonuclease 1 (APE1), aided by Poly [ADP-ribose] polymerase 1 (PARP1), cleaves the DNA at the AP-site to generate a 3' OH and 5' deoxyribose phosphate (dRP) terminus. Then a deoxyribophosphodiesterase (dRPase) cleaves the dRP residue to generate a 5' phosphate at the single-stranded break (SSB). Finally, DNA polymerase and DNA ligase seal the nick in the DNA.<sup>49–51</sup>



**Figure 1.4. DNA Repair Pathways.** (a) Repair of non-bulky base lesion by base excision repair (BER). (b) Repair of bulky, helix-distorting lesions by nucleotide excision repair (NER). (c) Replacement of mismatched bases by mismatch repair (MMR). (d) Repair of DSBs by more accurate classical nonhomologous end-joining (cNHEJ) with minimum loss of DNA or backup pathway alternative nonhomologous end-joining (aNHEJ) with significant DNA loss. (e) DSB repair by precise RAD51-catalysed homologous recombination (HR) or error-prone single-strand annealing (SSA), which results in significant loss of DNA. Figure adapted from Nickoloff, 2015.<sup>29</sup>

### *Nucleotide Excision Repair*

NER is responsible for removing bulky, helix-distorting lesions such as pyrimidine dimers and large DNA adducts, which may be induced, for example by UV light. Lesion recognition proteins are responsible for surveying the DNA for lesions and recruiting effector proteins. The unwinding of DNA around a lesion and stabilising of the resulting ssDNA requires a multitude of proteins including the transcription factor II H (TFIIH) complex, the Cdk-activating-kinase (CAK) complex, and XPs (Xeroderma pigmentosum, complementation proteins) named XPG and XPA. As the ssDNA is produced it is coated with RPA which activates XPG and a second endonuclease, excision repair cross-complementing rodent repair deficiency, complementation group 1 (ERCC1)-XPF, which create SSBs approximately 15 nucleotides either side of the lesion. A helicase then removes the oligonucleotide containing the lesion before finally DNA polymerase fills the single-strand gap and the ends are ligated by DNA ligase 3 (LIG3), XRCC1 (X-ray repair cross-complementing protein 1) and DNA ligase 1 (LIG1).<sup>52</sup>

### *Mismatch Repair*

In MMR there is no damage to the nucleotides but by some means mismatched bases exist on otherwise complementary DNA strands. Such mismatches can arise due to replication errors, strand exchange during HR or deamination of 5'-methyl cytosine to produce thymidine, which forms a G-T mismatch. MMR can also be utilised to remove loops that arise when bases are inserted or deleted, which again can arise due to replication errors or during HR. MMR features two families of proteins; heterodimeric homologues of bacterial MutS, MSH (MutS homologue), and MutL, MLH (MutL homologue). Repair is initiated by mismatch recognition by heterodimers MutS $\alpha$  (MSH2-MSH6) or MutS $\beta$  (MSH2-MSH3). These search for and interact with MutL $\alpha$  (MLH1-PMS2), PCNA (proliferating cell nuclear antigen) and RFC, to activate the endonuclease activity of PMS2 (post-meiotic segregation protein 2), which introduces a SSB. This generates an entry point for exonuclease 1 (EXO1) which removes the mismatch by single-strand excision. The DNA gap is then filled by DNA polymerases  $\delta$  or  $\epsilon$  (Pol $\delta$  or Pol $\epsilon$ ) and ligated by LIG1.<sup>29,53</sup>

BER, NER and MMR are reasonably accurate repair mechanisms as they are able to utilise the intact complementary DNA strand to direct repair of the lesion. Despite this, as the DNA is repaired it is vulnerable to localised mutagenesis. Repair polymerases are typically less accurate than replicative polymerases and the abasic intermediate produced during BER is susceptible to translesion DNA synthesis by Y-family polymerases, which are error-prone.<sup>29,54</sup> However, these challenges rather pale into insignificance compared to those posed by repair of DSBs.

### *Nonhomologous End-Joining*

NHEJ is composed of two subpathways, classical NHEJ (cNHEJ) and alternative NHEJ (aNHEJ). Of these two pathways, cNHEJ is the more accurate and, fortunately, the DSB repair pathway predominantly utilised in mammalian cells. It appears that aNHEJ is a backup pathway, generally used only where cells are defective in cNHEJ.<sup>55</sup> cNHEJ requires little to no end-resection. It is initiated by Ku70-Ku80 binding to dsDNA ends and recruiting and activating DNA-PKcs. Together Artemis, a nuclease responsible for processing of broken ends<sup>56</sup>, and Metnase, a nuclease and protein methylase that recruits other NHEJ factors<sup>57</sup>, promote association of the broken ends despite poor base pairing. At this stage DNA-PKcs dissociates from the DNA ends, which are then ligated by DNA ligase 4 (LIG4) with the aid of accessory factors XRCC4 (X-ray repair cross-complementing protein 4) and XLF (XRCC4-like factor).<sup>29,55</sup> In contrast, moderate end-resection is key to aNHEJ. aNHEJ is initiated by PARP1 binding to the DNA ends and recruiting MRN and CtIP (C-terminal-binding protein), which are responsible for exposing and aligning microhomologies at least five bases in length. The ends are then ligated by LIG3 and XRCC1 and the flaps trimmed by flap endonuclease 1 (FEN1).<sup>55,58</sup>

### *Homologous Recombination*

HR is also comprised of two pathways, one relatively precise pathway mediated by RAD51, and one more error-prone, RAD51-independent pathway, known as single-strand annealing (SSA). Both of these subpathways require extensive end-resection of hundreds or even thousands of bases. Perhaps it is understandable then that, as previously mentioned, the choice between DSB repair pathways seems to be regulated by proteins which control resection, such as 53BP1 and BRCA1.<sup>45</sup> The end-resection is conducted initially by MRN and CtIP, as in aNHEJ, and then by BLM helicase, DNA2 (DNA replication ATP-dependent helicase/nuclease) and EXO1 (Exonuclease 1).<sup>59</sup> This produces long 3' ssDNA tails coated with RPA, which, in the RAD51-dependent pathway, is exchanged for RAD51 by mediator proteins such as BRCA2 (breast cancer type 2 susceptibility protein) and RAD52. The RAD51-coated ssDNA searches for and invades homologous sequences and, once in position, is extended by DNA polymerase before release and capture by the DNA end on the opposite side of the DSB.<sup>29</sup> Where only one end invades, in synthesis-dependent strand annealing (SDSA), there is little chance of deleterious crossovers occurring. However, in some cases both ends invade, creating a double Holliday junction that must be resolved without occurring crossovers.<sup>60</sup> Such crossovers can result in loss of heterozygosity (LOH), known as gene conversion. To reduce the likelihood that crossovers cause large-scale genome rearrangements RAD51-dependent HR is typically restricted to S and G2 phases when sister chromatids, the most accurate repair templates, are in close proximity.<sup>29</sup> The SSA pathway is independent of RAD51

activity and is believed to depend instead on RAD52.<sup>60</sup> Extensive end-resection is utilised to expose complementary sequences in direct repeats. When uncovered these sequences are annealed to form a deletion product which lacks one of the repeats and any intervening sequence. SSA is an efficient method of repair for regions containing closely spaced repeats but otherwise results in significant loss of DNA.<sup>29</sup>

Accurately repairing DSBs is more challenging than SSBs. NHEJ is frequently inaccurate, generating short deletion or insertion mutations of up to 20 nucleotides. In addition, if two DSBs occur simultaneously on different chromosomes NHEJ can enable translocations. HR, particularly SDSA, is typically more accurate, but because any two homologous sequences can be engaged, it is associated with a significant risk of genome rearrangements. This might include deletions, inversions, amplifications, LOH or translocations.<sup>29</sup>

## 1.2 DNA Repair Dysregulation in Cancer

DNA damage is ubiquitous and clearly DDR plays a critical role in the accurate transmission of genetic information to daughter cells. However, as seen, DNA repair is primarily engineered to restore the chemical integrity of DNA with little regard for genetic integrity. As a result, DNA damage can generate mutations, which can range wildly in extent from point mutations, trinucleotide repeat expansions and contractions, gene duplications, deletions and inversions, to large-scale chromosomal rearrangements. This plays an important role in development of the immune system and evolution but can also be detrimental and ultimately lead to cancers and other genetic diseases.<sup>29</sup>

### 1.2.1 *Hallmarks of Cancer*

Hanahan and Weinberg describe eight hallmarks of cancer cells which govern the transformation of normal cells into malignant cancers; six well established and two recently emerging.<sup>61,62</sup> It is suggested that normal cells evolve progressively to become tumourigenic and ultimately malignant by successive acquisition of these hallmark capabilities. The proposed capabilities include both traits of the cancer cells and contributions from the tumour microenvironment. They include: sustaining proliferative signalling, evading growth suppressors, avoiding immune destruction, enabling replicative immortality, activating invasion and metastasis, inducing angiogenesis, resisting cell death and deregulating cellular energetics, Figure 1.5.<sup>62</sup>

#### *Sustaining Proliferative Signalling*

Arguably the principal trait of cancer cells is their aptitude for sustained proliferation. In healthy tissues, cell growth and division is carefully regulated by production and release of growth

promoting signals. These maintain a constant cell number and preserve normal tissue architecture and function.<sup>63</sup> In cancer cells, these signals are deregulated in a number of ways such that they are able to grow and divide more frequently. Firstly, the cells may themselves produce growth factor ligands to which they can respond, affording autocrine stimulation of proliferation. Alternatively the cells may signal to the normal cells in the tumour-associated stroma, which respond by supplying growth factors.<sup>64</sup> Some cancer cells display an increased number of receptor proteins at their cell surface making them hyperresponsive to any small amount of growth factor ligand, and others bear receptor molecules which are structurally altered so as to facilitate ligand-independent firing.<sup>62</sup> Furthermore, constitutive activation of signalling pathway components downstream of these receptors can relieve the necessity for any activation of the receptors. Approximately 40% of human melanomas harbour mutated BRAF. The most common mutation, occurring in 89% of cases, is the V600E substitution, which increases the catalytic activity of BRAF. This affords constitutive signalling to the mitogen-activated protein kinase (MAPK) pathway, which should otherwise only be triggered by activated growth factor receptors.<sup>65</sup> Finally defects can exist in feedback mechanisms that would normally dampen signalling to ensure homeostasis. One example involves PTEN (phosphatase and tensin homologue) phosphatase, which typically counteracts the activity of PI3K (phosphatidylinositol 3-kinase) by degrading its product. Loss-of-function mutations in PTEN therefore result in amplification of PI3K signalling and promote tumourigenesis.<sup>66</sup>

#### *Evading Growth Suppressors*

As well as continually inducing growth, cancer cells must evade those programs designed to suppress it. These programs are typically dependent on tumour suppressor genes such as RB (Retinoblastoma-associated) and p53.<sup>67</sup> RB is responsible for transducing primarily extracellular signals and decides whether a cell should proceed to growth and division. Cancer cells with defects in RB pathway function are permitted to proliferate continuously. P53 is responsible for halting cell cycle progression if the genome has suffered excessive damage or if levels of glucose, oxygenation, nucleotide pools or growth-promoting signals are abnormal. If irreparable damage has been sustained it can also induce apoptosis. In the absence of competent p53, cancer cells are again able to proliferate persistently. In addition, normal cells are prevented from excessive proliferation by 'contact inhibition', whereby the cell-to-cell contacts formed in dense populations suppress further proliferation. Moesin-ezrin-radixin-like protein (Merlin), the Neurofibromin-2 (NF2) gene product, coordinates contact inhibition, by coupling cell-surface adhesion molecules to transmembrane receptor tyrosine kinases. This stabilises the junctions and restrains kinase activity, and sequesters growth factor receptors. In this way Merlin limits



the cells ability to emit mitogenic signals.<sup>68</sup> Loss of Merlin has been shown to trigger neurofibromatosis.<sup>62</sup>



**Figure 1.5. The Hallmarks of Cancer.** Illustration of the eight proposed hallmark capabilities of cancer cells. Six are well established but the deregulation of cellular energetics and avoidance of immune destruction are more recently emerging. Figure adapted from Hanahan & Weinberg, 2011.<sup>62</sup>

#### *Resisting Cell Death*

It is well established that programmed cell death acts as a natural barrier to cancer development. Apoptosis is initiated in response to many of the physiological stresses which cancer cells experience in tumourigenesis or during treatment, such as signalling imbalances, or DNA damage due to rapid proliferation. Perhaps it is therefore unsurprising that apoptosis is attenuated in tumours which progress to malignancy.<sup>69</sup> The apoptotic regulators can be divided into two circuits, the extrinsic and intrinsic apoptotic programs. The extrinsic pathway involves cell-surface 'death receptors' such as FAS and TNFR (tumour necrosis factor receptor). When ligated these form 'death-inducing signalling complexes', which activate caspase-8 and thereby initiate a proteolysis cascade that results in apoptosis. The intrinsic pathway is the primary programme responsive to survival factor signals, cell stress and injury and is therefore more extensively implicated in deterring cancer pathogenesis.<sup>62</sup> The central locus is the mitochondrion, which stores pro-apoptotic effectors in its intermembrane space. The permeability of the mitochondrion is determined by the balance between pro- and anti-apoptotic members of the Bcl-2 (B-cell lymphoma-2) protein family. Bcl-2 and its closest relatives inhibit apoptosis by binding to and suppressing Bax (Bcl-2 Associated X) and Bak (Bcl-2 homologous antagonist/killer) at the outer mitochondrial membrane. Damage signals are transduced by BH3-only proteins (Bcl-2 family members with a single Bcl-2 homology-3 domain)

which interact with Bcl-2 and the other pro-survival proteins. This alleviates inhibition of Bax and Bak and initiates release of pro-apoptotic signalling proteins from the mitochondrion. One such factor is cytochrome c which triggers activation of caspase-9 and initiates the proteolysis cascade that results in apoptosis.<sup>70</sup> P53 induces apoptosis in response to severe DNA damage and chromosomal abnormalities by upregulating expression of BH3-only proteins.<sup>71</sup> Tumour cells can limit or escape apoptosis by a variety of strategies including loss of function of p53, increased expression of anti-apoptotic regulators or decreased expression of pro-apoptotic regulators.

Not all cell death is by apoptosis however. In contrast to apoptotic cells, which contract and are consumed, necrotic cells become bloated and explode, releasing their contents into the surrounding tissue. This releases proinflammatory signals and recruits inflammatory cells, which can be tumour promoting as they can enable angiogenesis, proliferation and invasiveness. Consequently, metastatic tumours may benefit from tolerating some necrotic cell death so as to recruit these inflammatory cells.<sup>62,72</sup>

#### *Enabling Replicative Immortality*

Normal cells are only able to journey through a limited number of cell growth-and-division cycles before induction of senescence or cell death. Cancer cells however exhibit unlimited replicative potential, termed immortality. It is understood that this ability to replicate indefinitely is associated with telomeres. Telomeres are composed of tandem TTAGGG repeats, which cap the ends of chromosomes to protect the DNA from recombination and degradation and maintain cell viability. They are progressively shortened at each cell division due to incomplete replication of linear chromosomes, the 'end-replication problem', and thereby act as a molecular clock, which underlies ageing. Telomerase, the DNA polymerase responsible for extension of telomeres, is virtually absent in normal cells but is upregulated to functionally significant levels in approximately 90% of immortalised cells. The extension of telomeric DNA counteracts the progressive erosion and is correlated with resistance to senescence and apoptosis.<sup>73</sup> Furthermore, tumour cells must pass through a significant number of successive telomere-shortening cell divisions prior to successful upregulation of telomerase. This delayed acquisition of telomerase function is conducive to generation of tumour-promoting mutations.<sup>74</sup>

#### *Inducing Angiogenesis*

Like any tissue, tumours need to receive nutrients and oxygen and expel metabolic waste and carbon dioxide. To achieve this they induce angiogenesis to develop neovasculature. In normal tissues vasculature is largely quiescent after embryogenesis. However, during tumour

progression angiogenesis is almost permanently activated, causing vasculature to grow new capillaries that help to sustain the growth.<sup>75</sup> Angiogenesis is induced early in the development of invasive cancers, substantiating its status as a hallmark of cancer.<sup>62</sup> It is understood that, like apoptosis, activation of angiogenesis is regulated by a balance of pro- and anti-angiogenic factors. For example, VEGF-A (Vascular Endothelial Growth Factor-A) is a pro-angiogenic gene that encodes ligands for tyrosine kinase receptors (VEGFRs), which orchestrate the growth of new blood vessels.<sup>76</sup> Contrastingly, thrombospondin-1 (TSP1) inhibits angiogenesis via direct effects on endothelial cell migration and survival and by sequestering and restricting the movement of VEGF. Upregulation of pro-angiogenic genes or downregulation of anti-angiogenic genes can induce angiogenesis.<sup>77</sup> Finally, as eluded to earlier, it appears that inflammatory cells of the innate immune system are recruited to developing tumours to help to initiate and sustain angiogenesis.<sup>78</sup> The extent of variation of neovascularisation in different tumours is illustrative of the complexity of control. Interestingly some oncogenes that drive proliferative signalling can also induce angiogenesis, highlighting the potential that hallmark capabilities can be coregulated.<sup>62</sup>

#### *Activating Invasion & Metastasis*

Fundamental also to malignant cancers is their ability to invade and metastasise. To complete these processes the tumour cells must be able to detach, migrate, gain access to and from the blood or lymphatic vessels, micrometastasise at a new location and finally colonise to form a macroscopic tumour. Three distinct modes of cancer cell invasion have so far been identified; mesenchymal, collective invasion and amoeboid. Of these the best understood is mesenchymal invasion, which is regulated by the epithelial-mesenchymal transition (EMT) program.<sup>62</sup> In EMT a subset of tumour cells switch off epithelial markers such as E-cadherin, a key cell-to-cell adhesion molecule, and switch on mesenchymal markers such as vimentin. This leads to loss of cell polarity, cytoskeleton reorganisation, loss of adherens junctions, expression of matrix degrading enzymes, increased motility and heightened resistance to apoptosis.<sup>79</sup> The primary regulator of EMT is transforming growth factor- $\beta$  (TGF- $\beta$ ).<sup>80</sup> TGF- $\beta$  functions through a collection of transcription factors including Snail, Slug, Twist and Zinc finger E-box-binding homeobox 1/2 (ZEB1/2), which coordinate the EMT and migratory processes required during embryogenesis and have been found to be expressed in various combinations in malignant tumours.<sup>81</sup> There is increasing evidence that the acquired capability for invasive growth originates from the tumour-associated stromal cells neighbouring the cancer cells.<sup>82,83</sup> Interestingly, if this proves true it is plausible that tumours could metastasise without the requirement for mutations other than those for primary tumour formation.<sup>62</sup> The process of colonisation is highly complex, varying in

different tissue types, and it is unclear when and where tumour cells develop the ability to colonise foreign tissues. The capability may arise fortuitously during primary tumour formation or may develop only in response to selective pressures at the new site. Either way it is highly likely that colonisation requires the establishment of a permissive tumour environment including stromal support cells.<sup>62</sup>

#### *Deregulating Cellular Energetics*

The perpetual cell proliferation associated with cancer cells must be fuelled by adjusted energy metabolism. In aerobic conditions normal cells harvest ATP by processing glucose first to pyruvate by glycolysis and then to carbon dioxide by oxidative phosphorylation in the mitochondria. In anaerobic conditions little pyruvate can be processed by the mitochondria so energy metabolism is limited largely to glycolysis. As first observed by Warburg<sup>84</sup>, cancer cells often reprogram their glucose metabolism to utilise glycolysis almost exclusively, independent of the presence of oxygen. Considering that 'aerobic glycolysis' is an inefficient way to generate ATP this is somewhat surprising. However, it is suggested that the abundance of glycolytic intermediates produced can be utilised in generating biomass including nucleosides, amino acids or lipids, which enables the biosynthesis of macromolecules and organelles required to produce a new cell.<sup>85,86</sup> This increased reliance on glycolysis is associated with activated oncogenes and downregulated tumour suppressors as well as being accentuated in some tumours by hypoxic conditions.<sup>87</sup> As deregulation of cellular energetics seems to be as prevalent as other hallmarks of cancer it has been designated as an emerging hallmark.<sup>62</sup>

#### *Avoiding Immune Destruction*

It is believed that the immune system acts as a barrier to tumour formation and progression. Increased tumour growth has been observed in immunodeficient mice, for example with deficiencies in development or functioning of natural killer (NK) cells, compared to immunocompetent controls.<sup>88</sup> Transplantation experiments have also shown that cancer cells that originally developed in immunodeficient mice are unable to establish tumours in immunocompetent mice, whereas cells from the immunocompetent mice can initiate tumour growth in any new host.<sup>88</sup> There is increasing support for the existence of a similar immunosurveillance mechanism in human cancers.<sup>89</sup> It is suggested that cancer cells may avoid immune destruction by disabling components of the immune system destined to eliminate them.<sup>62</sup> The cancer cells may for example secrete TGF- $\beta$  or other immunosuppressive factors.<sup>90</sup> In light of this research, avoiding immune destruction is identified as a second emerging hallmark of cancer.<sup>62</sup>

Concordantly, over the past decade there has been a surge of interest in the field of cancer immunotherapy, which aims to treat cancer by generating or augmenting an immune response. In clinical trials immunotherapies have consistently improved the overall survival of patients with advanced-stage cancers and three immune-checkpoint-blocking monoclonal antibodies have already been approved for treatment of several cancer types in the USA.<sup>91</sup>

### 1.2.2 *Genome Instability*

The DDR operates very effectively, and despite the very high occurrence of DNA lesions, mutation rates in normal mammalian cells are low, at approximately  $10^{-10}$  per base per cell generation, or  $10^{-6}$ - $10^{-8}$  per gene per cell generation.<sup>92</sup> It is estimated that converting a normal cell to a metastatic cancer cell requires 3-10 mutations in key growth cells and tissue regulatory genes.<sup>29</sup> If mutations arise independently, the chance of accumulating five in a single cell is phenomenally small,  $10^{-30}$ - $10^{-40}$ . However, unfortunately mutations do not always arise independently.

Early insights into the role of the genome in cancer development emerged in the late nineteenth and early twentieth centuries when David von Hansemann and Theodor Boveri observed chromosomal aberrations in dividing cancer cells.<sup>93,94</sup> As discussed above, cancers arise when cells obtain hallmarks such as altered cell growth properties, immortality and defective programmed cell death pathways. Progression to the more aggressive metastatic state requires enhancements such as angiogenesis, tissue invasion and adaptability to new environments.<sup>29</sup> Despite it having been well established that cancer is linked with genome instability, it was difficult to ascertain whether this was causative or a secondary manifestation.<sup>95</sup> There are now several studies however in which it has been definitively shown that genome instability precedes cancer.<sup>96,97</sup> Although the genetic changes responsible for initiating tumourigenesis and promoting tumour progression differ in different cancer types, the early acquisition of defects in DNA repair is common. Such DNA repair defects cause genetic instability which considerably increase the rate at which mutations are acquired, including gain-of-function mutations in proto-oncogenes and loss-of-function mutations in tumour suppressor genes.<sup>29</sup>

As well as initiating tumourigenesis, the increased mutagenesis afforded by acquisition of a DNA repair defect enables evolution of the cancer cells to promote tumour progression. The fitness of tumour cells is determined by the balance of the beneficial effects of mutational variation, which facilitate adaptation to changing environmental pressures, and any detrimental effects of mutation. The limit of the cellular capacity to tolerate further mutagenesis is termed the 'error catastrophe' limit.<sup>98</sup> Providing that mutation rates remain below this limit, the tumour cells are

able to experiment with successive mutations that may afford an improved response to environmental stresses or resistance to therapy. This has two interesting repercussions. Firstly, the majority of the tens of thousands of mutations present in every cancer genome are not believed to have been selected for or be causally linked to the pathogenesis of the cancer and are instead 'passenger mutations'.<sup>98,99</sup> Secondly, some mutations acquired late in the evolution of the cancer cell may only be advantageous or even tolerated because of a preceding mutation, establishing oncogene 'addiction'.<sup>100</sup> Most cancer patients are treated with chemotherapeutics or ionising radiation that induce DNA damage. These approaches were developed to exploit the high proliferation rate of tumour cells compared to healthy human cells, as cells actively replicating DNA are more susceptible to the cytotoxic effects of DNA damage.<sup>101</sup> However, perhaps their anticancer effects may also be attributed in part to inducing increased mutational load on the cancer cells which pushes them over the error catastrophe limit.<sup>98</sup>

Unfortunately, tumour cells can exhibit resistance to traditional therapies. In some cases this may reflect a poor tailoring of treatment to the specific DNA defect, whereby the selected therapeutic induces DNA damage repaired by pathways still functional in that particular tumour. However, resistance may also reflect upregulation of other DNA repair pathways or defects in apoptosis and other programmed cell death. For example, approximately 50% of tumours carry defects in p53.<sup>29</sup> Furthermore, the rapid evolution of tumour cells afforded by genome instability allows them to adapt and acquire resistance to therapies. Traditional therapies are effective in killing tumour cells in bulk but the DNA damage they induce can spawn and ultimately select for cancer cells that are resistant to therapy and potentially more aggressive, which can result in tumour recurrence and progression to a metastatic state.<sup>29</sup> As such, despite short-term benefits, the therapies do not necessarily increase long-term patient survival. It seems that current therapeutic strategies require re-evaluation. To improve patient survival, new therapeutics must be able to destroy the entire population of tumour cells. This is not easily achieved alongside minimising the effects on normal tissue. Traditional cancer therapies also cause many serious side effects including neurological and gastrointestinal problems, fatigue, fever and liver and kidney failure, and additional late effects including cardiac disease, infertility, hearing loss and secondary tumours.<sup>29,102</sup> It is hoped that therapeutics designed to selectively target tumour weaknesses could eradicate tumour cells while being less harmful to normal cells.

### 1.3 Targeted Cancer Therapy

The predominant obstacle in developing cancer therapeutics is not finding chemicals that are able to kill cancer cells but identifying those which are capable of doing so at concentrations that spare normal tissue. Indeed, most available chemotherapeutic agents have surprisingly low

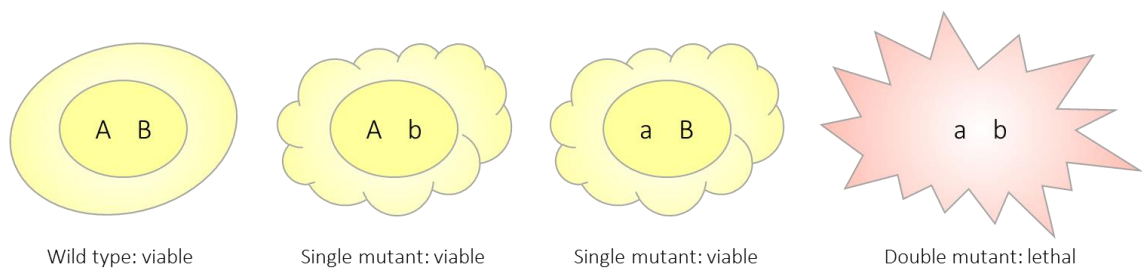
therapeutic indices, defined as the dose at which toxic effects are observed divided by the dose required for the therapeutic effect.<sup>100,103</sup> This can result from a number of contributing factors relating to both the on- and off-target effects of the drugs, including the selectivity profile of the drug or the distribution and function of the protein target.<sup>100</sup> The tolerance level of normal tissues thereby limits the dose at which the drug can be delivered and increases the likelihood that some tumour cells will survive after treatment.<sup>29</sup> One particular challenge in developing chemotherapeutics is identifying a suitable target that affords the opportunity to preferentially target cancer cells over normal cells with any considerable degree of selectivity.

Two avenues can be explored in attempt to improve the therapeutic index of drugs. The first, the target-driven therapeutic index, requires identification of a target that is specific to the diseased tissue. Antibacterial agents, for example, are invariably safe because they target proteins present only in the invading organism and not in normal human cells.<sup>100</sup> This is not impossible for cancer therapeutics, especially as the number of oncogenes described in the literature continues to grow.<sup>104</sup> For example, Imatinib (Gleevec) and later generation tyrosine kinase inhibitors have been used to good effect to treat chronic myelogenous leukaemia by inhibition of tumour-specific BCR-ABL (breakpoint cluster region-Abelson murine leukaemia viral oncogene homologue).<sup>105</sup> However, this approach is not always applicable. Suitable targets are limited as many of the genetic alterations frequently observed in tumours, particularly those of DDR proteins, are loss-of-function mutations; also it can be difficult to develop drugs that specifically inhibit oncoproteins without affecting their normal counterparts, as these often only result from point mutations.<sup>100</sup> Additionally, such highly specific therapies may not be effective in targeting genetically heterogeneous tumours.<sup>29</sup> The second, context-driven therapeutic index, is dependent on an enhanced requirement for the target protein in the diseased tissue compared to normal tissue.<sup>103</sup> This approach is more broadly applicable as it enables targeting of cancer cells with loss-of-function mutations, which are traditionally harder to treat, because it is technically difficult to recapitulate tumour suppressor function. It is clear that these contextual differences are therapeutically exploitable from the wealth of effective anticancer drugs already in use that target proteins common to normal and cancer cells.<sup>100</sup> However, it must be acknowledged that this approach will generate drugs vulnerable not just to toxicity via off-target effects but also potentially deleterious on-target effects.

### 1.3.1 *Synthetic Lethality*

One key approach in developing deletion-guided chemotherapies is to exploit synthetic lethal interactions. The concept of synthetic lethality was first defined by Dobzhansky in relation to drosophila genetics. He identified genes which separately were not lethal to homozygotes but

became lethal when combined by crossing over.<sup>106</sup> Correspondingly, two mutations are now termed ‘synthetic lethal’ if neither alone affects the viability of the cell but in combination they result in cell death, Figure 1.6. Mutations are termed ‘synthetic sick’ if they impair cellular fitness more in combination than individually, but this impairment is non-lethal. It has now been 20 years since it was first suggested that synthetic lethality could be utilised in anticancer drug discovery.<sup>107</sup> Protein products of genes which are synthetic lethal with mutations commonly observed in cancers should represent promising targets for chemotherapeutics, provided that they are amenable to pharmacological regulation. Ideally, if genes A and B are synthetic lethal, inhibition of B should kill cancers cells with mutant A with no effect on normal cells. However, in practice it is more common that inhibitors of B either exhibit some toxicity in normal cells, or are non-lethal in cancer cells, because A and B fall short of being truly synthetic lethal.<sup>108</sup> This will limit the specificity of therapeutics targeting B, but it is hoped that these inhibitors could still afford a significant therapeutic index.<sup>100</sup>



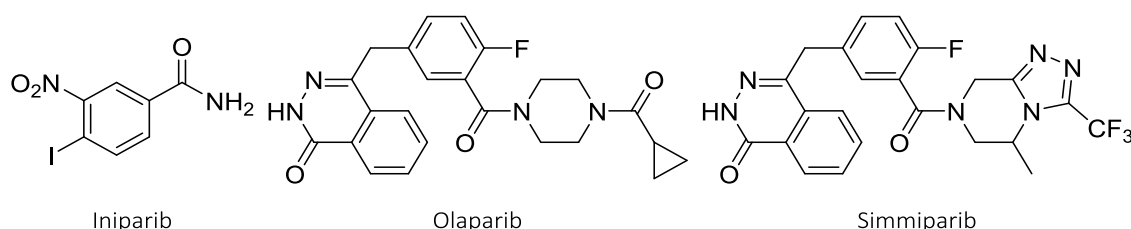
**Figure 1.6. Synthetic Lethal Interactions for Two Genes.** Two genes, A and B, are said to be synthetic lethal if mutation of neither alone affects the viability of the cell but simultaneous mutation of both genes results in cell death. Lower case letters denote mutants.

The best renowned example of a synthetic lethal interaction in human cancer is that of BRCA1 or BRCA2 with PARP1. In 2005, the Helleday and Ashworth labs independently showed that BRCA1- or BRCA2-defective breast cancers were sensitive to PARP1 inhibitors.<sup>109,110</sup> Mutations in either BRCA1 or BRCA2 are associated with a predisposition to breast and ovarian cancers.<sup>111</sup> As discussed earlier, these genes are important in double-stranded break repair by homologous recombination. PARP1 is also involved in DNA repair, particularly the repair of single-stranded breaks. It appears that loss of function of PARP1 results in a hyper-recombination phenotype and collapse of replication forks. Loss of PARP1 correlates with increased levels of RAD51 foci, where RAD51 is gathered at sites of DNA damage, and sister chromatid exchange, in which sister chromatids exchange DNA by breaking and rejoining with each other.<sup>112</sup> This suggests that the loss of PARP1 serves to increase the formation of DNA lesions to be repaired by HR. It was hypothesised and shown that inhibition of PARP1 in BRCA1- or BRCA2-defective cells, which are defective in HR, results in irreparable DNA lesions which cause chromatid aberrations and loss of viability.<sup>109,110</sup> These findings sparked great excitement. PARP1 inhibitors were capable of



eradicating BRCA-defective tumours in mice models. Furthermore, BRCA-defective cells showed incredible sensitivity to PARP1 inhibition in comparison to wild type cells, displaying a 50- to 1000-fold enhanced effect.<sup>110</sup> The PARP1 inhibitors were unique among cancer therapies exploiting DNA damage sensitivity in tumour cells as they displayed an impressive therapeutic index without the need to apply exogenous DNA damaging agents. The BRCA-PARP1 interaction is truly synthetic lethal and not synthetic sick, or synthetic sensitive.<sup>29</sup> It is uncertain how common such strong genetic interactions will prove to be in human cells.<sup>108</sup>

Clinical trials were quickly initiated for several PARP inhibitors, including Sanofi's iniparib and AstraZeneca's olaparib, Figure 1.7. Unfortunately this very promising class of compounds suffered a near-fatal blow when both iniparib and olaparib failed to demonstrate good efficacy in trials in patients with triple-negative breast cancer.<sup>113</sup> However, it seems that these failures were alarmingly easily explained. Despite PARP inhibitors having been clearly shown to confer synthetic lethality in cancers with BRCA1 and BRCA2 mutations, the trials utilised a larger, heterogeneous population including patients with intact BRCA genes as long as they displayed other homologous recombination defects. It later emerged that iniparib is also not a *bona fide* PARP inhibitor.<sup>113</sup> This illustrates the immense importance of good communication in drug development and regrettably delayed the approval of drugs.<sup>114</sup> Despite these setbacks, by 2016, 13 PARP inhibitors had been or were being tested in the clinic for cancer therapy and, after restarting of phase III clinical trials in 2003, olaparib was finally approved for the treatment of advanced ovarian cancer in patients with BRCA mutations by the European Union and the FDA in 2014.<sup>115</sup> Very recently another PARP inhibitor, simmiparib also entered clinical trials.<sup>116</sup> Unfortunately the battle is not yet won. Examples are emerging of BRCA-defective tumours which are gaining resistance to PARP inhibitors by mechanisms such as loss of PARP1 or reactivation of HR.<sup>117</sup>



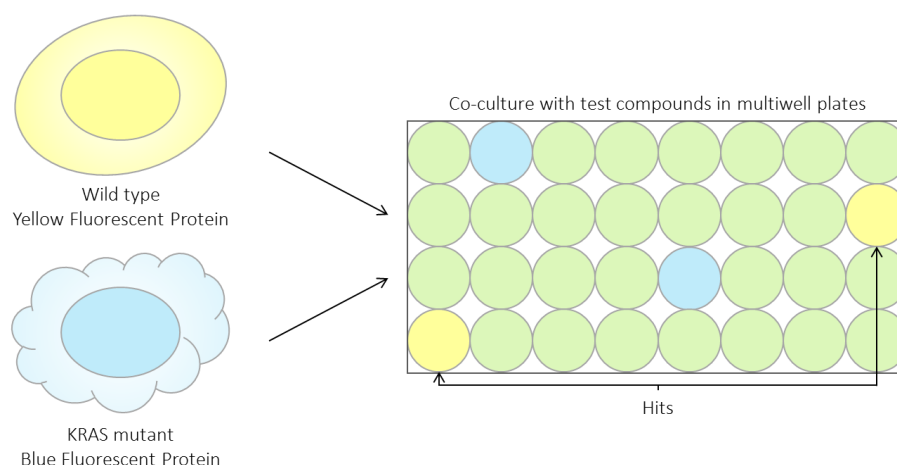
**Figure 1.7. Chemical Structures of Proposed PARP Inhibitors.** Iniparib, eventually proven not to be a functional PARP inhibitor, olaparib and simmiparib.<sup>115,116</sup>

Synthetic lethality can result both from two mutations having additive negative effects on a single essential biological pathway or, as just discussed, inactivating different but functionally overlapping pathways.<sup>107</sup> As a result, the web of synthetic lethal relationships is complicated and chaotic, with perhaps ten interactions per gene.<sup>100</sup> Consequently, despite the wealth of genetic

and epigenetic data available about cancers, molecular networks are not sufficiently well understood to enable accurate prediction of synthetic lethal relationships. For this reason screens have been developed, utilising chemical or genetic methodologies, to expose the synthetic lethal relationships of cancer-relevant mutations, with the hope of identifying opportunities for future therapeutics. Chemical and genetic synthetic lethal screens are likely to afford distinct but complementary results as the effect of drug-induced inhibition of enzymatic function may differ significantly from complete loss of expression of the protein.<sup>118</sup> For example, a drug might interfere with only one function of a multifunctional protein, may upregulate or downregulate protein function or may exert an effect via multiple protein targets.<sup>119</sup>

### 1.3.2 Chemical Synthetic Lethal Screens

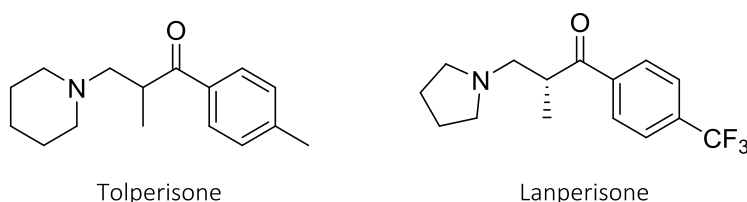
Chemical-based screens aim to directly identify molecules which induce a synthetic lethal response with cancer-relevant mutations in cell-based phenotypic assays. This screening technique benefits from preserving the cellular context of the protein function, averting any problematic disparities between the effect of a drug compared to complete loss of gene expression, and can provide a shortcut to utilise drugs already in clinical use. The opportunity and applications for such screens were highlighted in 1997 by Hartwell and Friend<sup>107</sup>, who quickly followed this with a screen of FDA-approved anticancer drug sensitivities in yeast deletion mutants defective in DNA repair or cell cycle checkpoint functions.<sup>120</sup>



**Figure 1.8. Fluorescence Based Synthetic Lethal Assay.** Isogenic cell-line pairs with and without KRAS mutations and expressing blue and yellow fluorescent protein respectively are grown in multiwell plates with drugs administered. A 'hit' is identified as a chemical which selectively kills the cells with mutant KRAS, creating a yellow well.<sup>121</sup> Figure adapted from Kaelin, 2005.<sup>100</sup>

The concept can be extended to human cells to identify compounds able to selectively inhibit cells with DDR deficiencies utilising isogenic human cell-line pairs grown in multiwell plates. This has been used to good effect in identifying compounds which afford selective killing of KRAS mutant cancer cells. Torrance *et al.* co-cultured colon cancer cells with mutant KRAS, which they

had engineered to express blue fluorescent protein, and wild type cells with the mutant KRAS deleted, engineered to express yellow fluorescent protein. A different compound was added to each well and differential killing was monitored using the ratios of blue to yellow fluorescence, Figure 1.8.<sup>121</sup> Several novel compounds including triphenyl tetrazolium and a cytidine nucleoside were identified which selectively killed the KRAS mutant cell line. In a more recent example, in 2011, Shaw *et al.* screened over 50,000 compounds using an ATP-based cell viability assay.<sup>122</sup> Two compounds showed potent and selective activity against the mouse embryonic fibroblasts expressing oncogenic KRAS compared to wild type controls. Of these, tolperisone was selected as the subject for further study and ultimately the derivative lanperisone was selected as their most potent hit, with an  $IC_{50}$  of 4  $\mu$ M, Figure 1.9. It was ascertained that lanperisone induces its cytotoxic effect by triggering oxidative stress; however, the precise target of the compound remains unclear. Indeed successful target identification can often be a problem related to chemical-based synthetic lethality screening.<sup>100</sup>



**Figure 1.9. Chemical Structures of Tolperisone and Lanperisone.**<sup>122</sup>

Such screens are particularly powerful and unique in their ability to correctly prioritise hits. It is not uncommon in HTS for approximately 1% of the compounds in a chemical library to inhibit the growth of cancer cells, which, from a screen of  $10^5$ - $10^6$  compounds, results in thousands of hits. In the absence of an appropriate filter, such as this specific killing of cells with cancer-relevant mutations, these hits are prioritised based on features such as potency, ease of synthesis, issues of intellectual property and ADME (absorption, distribution, metabolism and excretion) properties, which although are important, do not pertain to selectivity. This can potentially cause compounds which would offer selectivity to be missed.<sup>100</sup>

As alluded to, the frustration in chemical-based screens is that the precise protein targets and mechanism of action of the molecules responsible for the observed phenotype are unknown.<sup>123</sup> Successful target identification can often be difficult, especially if the drugs screened have broad activity profiles.<sup>100</sup> Three distinct and complementary approaches are typically utilised. Firstly, direct biochemical methods, employing biochemical affinity purification to find those target proteins which bind to the small molecules of interest. Secondly, genetic interaction and genomics methods, in which target hypotheses are generated based on knowledge and understanding of genetic interactions and modifiers (enhancers or suppressors) and tested using

gene knockout organisms, RNAi (RNA interference) or other small molecule modulators. Finally, computational inference methods, which rely on small molecule phenotypic profiling, ligand-based methods or structure-based methods.<sup>123</sup> Each of these approaches has strengths and limitations and typically labs will make the decision on which to employ based on their technical expertise. There are examples where the methods have been successfully implemented to determine the mechanism of action of a small molecule. An integrative chemical genomic and proteomic approach was employed to determine that K252a, a natural-product, potentiates neuregulin-1-dependent neuritogenesis by interaction with AAK1 (adaptor-associated kinase 1). Active and inactive analogues of the small molecule probe were utilised in stable isotope labelling by amino acids in cell culture (SILAC)-based affinity enrichment, and the resulting candidate targets were tested by RNAi-mediated gene silencing.<sup>124</sup>

There are however many instances in which the mechanisms of action have never been successfully unveiled. Additionally, the scope of chemical synthetic lethal screens is limited as only those targets for which there is already a tool inhibitor available can be investigated, which could result in superior targets being overlooked. Essentially therefore, complementary genetic synthetic lethal screens have also been developed.

### 1.3.3 Genetic Synthetic Lethal Screens

Genetic screens serve to identify genes which are synthetic lethal with the mutation of interest. In contrast to chemical screens, which are more diverse, in that the drugs employed can have a range of effects on their target protein or proteins, genetic screens can typically only be used to investigate downregulation of gene expression. However, they are invariably applicable and allow a broader scope to investigate genes whether or not there is an existing tool inhibitor.<sup>100</sup>

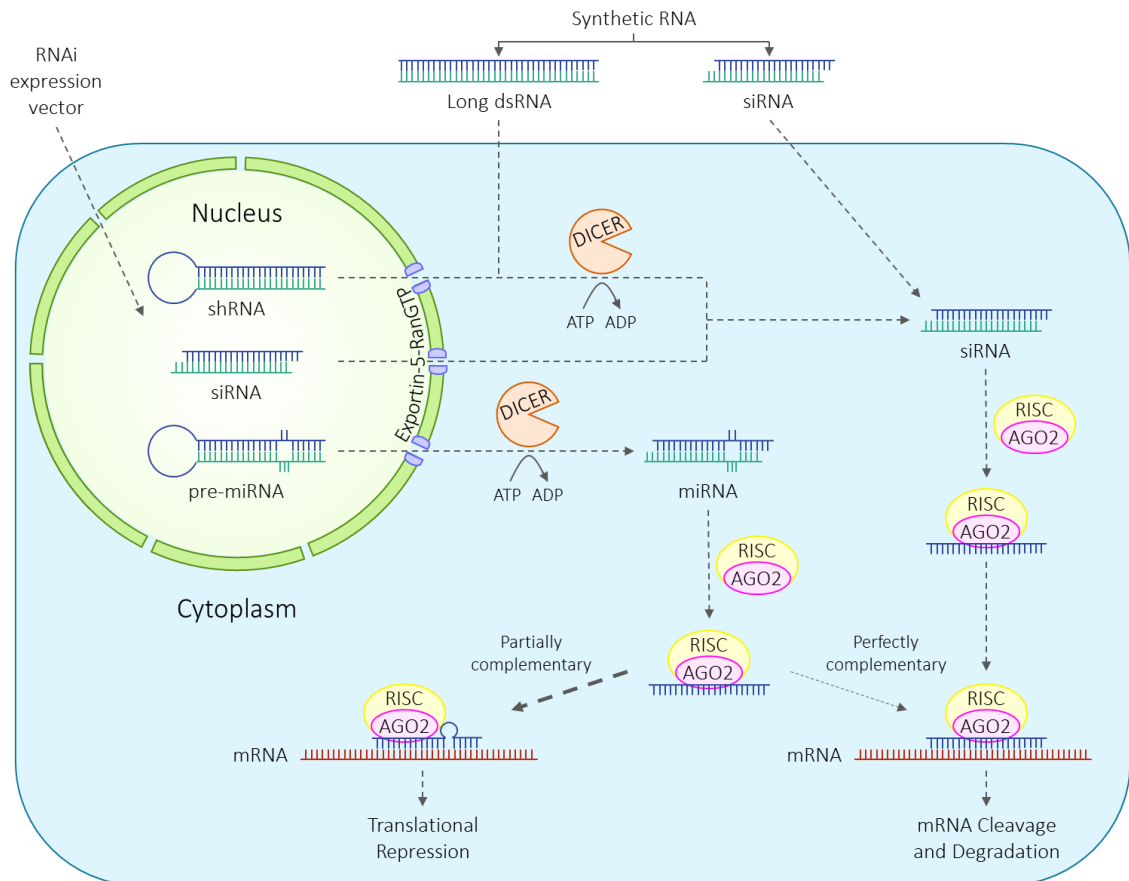
Traditionally genetic screens for synthetic lethal interactors were conducted in model organisms amenable to forward-genetic approaches, such as yeast and drosophila. Screens in yeast have been key in developing understanding of the fundamentals of synthetic lethal interactions, but they are limited, as many human cancer genes have no yeast orthologue. Forward-genetic screens in drosophila are more laborious but afford the advantage that their genomes contain orthologues of most human oncogenes and tumour-suppressor genes.<sup>100</sup> For example, inactivating mutations in the retinoblastoma (Rb) pathway are common in human tumours, so in 2005, Edgar *et al.* performed a clonal screen in the eyes of drosophila to identify mutations which selectively eliminate Rb mutant cells. Peptidyl prolyl isomerase was identified as synthetic lethal with Rb.<sup>125</sup>

Reverse-genetic approaches are now overtaking forward approaches. RNAi is an endogenous cellular process that enables targeted gene silencing. It was first observed in 1998 in *Caenorhabditis elegans* where double-stranded RNA (dsRNA) was uncovered as the potent trigger of gene silencing.<sup>126</sup> It is thought to have evolved to protect hosts from transposable elements and viruses that utilise dsRNA to propagate.<sup>127</sup> RNAi was quickly adopted as a research tool in invertebrate model systems where it facilitated functional genomic analysis.<sup>128,129</sup> Unfortunately the methodologies were limited in scope because in mammalian cells dsRNA tended to induce an antiviral interferon response which resulted in global shutdown of protein synthesis and therefore cell death.<sup>130</sup> Thankfully though, this obstacle was overcome when it was observed that during RNAi dsRNA is cleaved into smaller siRNA fragments and these are capable of triggering RNAi in mammalian cells without activating the antiviral response.<sup>131</sup>

RNAi is now known to be a widespread phenomenon, occurring in fungi, plants and animals, in all of which it is triggered by long dsRNAs.<sup>132</sup> All protein-encoding genes are transcribed in the nucleus by RNA polymerase II. The primary RNA transcript is processed by splicing to produce mature messenger RNA (mRNA), which is then exported from the nucleus into the cytoplasm. Here ribosomes catalyse translation of the mRNA to form proteins.<sup>133</sup> This is where RNAi exerts its silencing effect, Figure 1.10. RNAi is initiated by the interaction of long dsRNAs with DICER, an endoribonuclease that cuts the RNA into short segments, approximately 21 nucleotides long, to create siRNAs with 3' overhangs and 5' phosphate groups. Dicer can process linear dsRNAs and hairpin-RNA substrates.<sup>132</sup> The resulting siRNA binds to argonaute-2 (AGO2) and assembles into RNA-induced silencing complex (RISC).<sup>134</sup> The two strands of the siRNA are sense and antisense with respect to the target mRNA. The antisense 'guide' strand acts as a template for the sequence-specific gene silencing. The sense strand is a 'passenger' and is degraded upon RISC assembly. The guide-strand-containing RISC binds to the target mRNA and silences gene expression by RNA cleavage and degradation. Only when there is perfect complementarity between the guide and target strands will the mRNA be cleaved. Therefore, RNAi affords reasonable specificity in silencing genes.<sup>132</sup> Gene silencing can also be provoked by endogenous, small non-coding RNAs called microRNAs (miRNAs). miRNAs are transcribed in the nucleus by RNA polymerase II to form primary miRNA transcripts, which are cropped to 70-nucleotide pre-miRNAs by the DROSHA-DGCR8 (DiGeorge syndrome critical region 8) complex.<sup>135</sup> These pre-miRNAs are moved to the cytoplasm by nuclear transport receptor complex, exportin-5-RanGTP<sup>136</sup>, where they are then processed by DICER to mature miRNAs. The miRNAs are subsequently loaded on to RISCs by AGO2 and guide gene silencing. Usually only part of the miRNA, known as the seed, pairs with the target mRNA. This imprecise matching allows miRNAs

to target hundreds of endogenous mRNAs. Without a perfect match between the miRNA and mRNA the target does not undergo cleavage and degradation but translation is repressed.<sup>132</sup>

Researchers have now been harnessing RNAi as a tool for sequence-specific gene silencing for over 15 years. siRNAs, designed for endogenous mRNA transcripts, are introduced to cells or whole organisms to afford targeted gene silencing. The siRNAs can be introduced either directly as synthetic siRNAs or siRNA precursors, such as short hairpin RNAs (shRNAs) and long dsRNAs, or by addition of plasmids expressing RNA, Figure 1.10.



**Figure 1.10. Mechanisms of RNAi.** RNAi is triggered by long dsRNA, siRNA, shRNA or miRNA. Pre-miRNA, dsRNA and shRNA require processing by DICER to form shorter siRNA duplexes approximately 21 nucleotides long. Processed siRNAs or miRNAs assemble with a collection of proteins including AGO2 to form RISC. During RISC assembly the sense 'passenger' strand of the RNA is degraded. The other antisense 'guide' strand is retained to produce activated RISC, which triggers sequence-specific mRNA degradation or translational repression depending on the extent of complementarity. Figure adapted from De Paula *et al.*, 2007.<sup>137</sup>

Ten years ago, experiments using individual RNAi reagents were being widely used in examining the functional roles of specific proteins and genes. For example the defining roles of KRAS, AURORA B kinase and plasminogen activator inhibitor 1 (PAI1) in tumourigenesis. Such approaches highlighted the huge potential for RNAi in drug discovery efforts and it was suggested that the development of genome-wide RNAi approaches could significantly reduce the cost and time involved in target identification and validation.<sup>138</sup> The development of genome-wide RNAi libraries, by academic and commercial groups, has made it possible to

conduct high-throughput genome-wide loss-of-function screens.<sup>139</sup> These libraries can be comprised of one or more types of RNAi reagents. Fundamentally, the library formats are divided into two types: collections of siRNA reagents, synthesised chemically or generated from complementary DNA (cDNA) by ribonuclease III digestion, or libraries of shRNA expression vectors, which can be viral or non-viral (plasmid). The siRNA libraries are typically employed in those situations where short-term gene silencing is sufficient. The siRNAs are not replicated by the mammalian cells and therefore become progressively diluted as the cells divide. The shRNA vector libraries provide long-term, stable gene silencing, as the vectors integrate into the genomic DNA and are therefore copied at cell division.<sup>138</sup> UP-TORR is an online tool which aims to curate accurate and up-to-date annotations of the available cell-based and *in vivo* RNAi reagents, to offer researchers an informed decision as to which are most appropriate for their experiments.<sup>140</sup>

Ideally, RNAi screening combines the powers of genetic screens and phenotypic assays. It can be adapted to identify genes or gene networks involved in a huge variety of biological processes by integration of appropriate assays, such as for signal transduction, cell viability, cell or organelle morphology, organelle or protein localisation and/or function, drug resistance or cellular response to pathogens.<sup>139</sup> Broadly speaking there are two typical formats for RNAi screening. The first, arrayed screening, aims to determine the effect of loss-of-function of individual genes on a phenotype of interest. This is achieved by transfecting with libraries of RNAi reagents separated one into each well of multiwell plates. Any suitable readout can be employed. Perhaps the simplest and most common is cell viability. The second approach utilises pooled shRNA vectors. Groups of library-derived vectors are introduced to cells and those cells are selected for a resulting phenotype of interest.<sup>138</sup> The target gene in these select cells is identified either by PCR (polymerase chain reaction) amplification and sequencing or by an approach known as barcode screening.<sup>141</sup>

Hundreds of large-scale cell-based and *in vivo* RNAi screens have now been carried out, which promise improved understanding of many complex cellular networks and processes.<sup>139</sup> However, this technology has proved to have a number of associated pitfalls, which have resulted in few of the genes identified in these screens being successfully validated. Chiefly, off-target effects have been linked to virtually every type of RNAi reagent.<sup>142</sup> This difficulty is exacerbated when scoring for inhibition of a complex phenotype regulated by many genes, such as cellular proliferation or viability.<sup>143</sup> Consequently, RNAi screening is afflicted with inter-screen hit discordance and poor reproducibility. In 2013, Bhinder and Djaballah conducted a systematic comparative analysis of screening data from 30 representative lethality-based studies.<sup>144</sup> They

noted a distinct lack of overlap between the studies, particularly those employing different RNAi technologies, with no genes common across the board. In addition, 80% of the genes identified by shRNA screening were found to be exclusive to the pooled format, raising doubt as to the merits of this approach.<sup>144</sup> It is suggested that gold standards and greater prudence in analysis of RNAi screening data are essential in improving the reproducibility of data.<sup>143,144</sup>

RNAi screening methodologies can be adapted to screen for synthetic lethal interactions. Theoretically, RNAis that exacerbate the mutant phenotype of tumour cells, or cells otherwise mutated in one particular gene, without affecting wild type cells, would be indicative of a synthetic lethal or synthetic sick interaction. This comparative approach may incidentally help to reduce false positives in screening that result from non-specific impairment to cellular fitness.

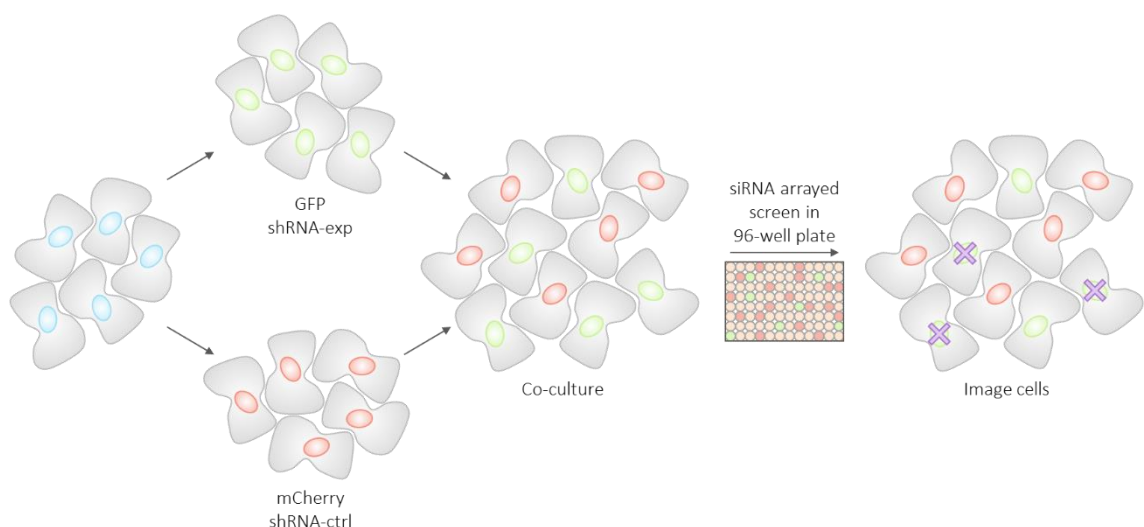
Combinatorial screens, in which two genes are simultaneously silenced, can be used to uncover synthetic lethal relationships between genes. For example, the Boone lab developed a screening method in *Saccharomyces cerevisiae* termed synthetic genetic array (SGA), which enabled them to cross a 'query' strain bearing one viable gene mutation with an array of ~5000 viable deletion mutants, select the resulting double mutants and score these for growth and fitness defects.<sup>145</sup> Those double mutants that exhibited greater impairment of fitness than expected from the combination of the corresponding individual mutations were indicative of synthetic lethal interactions. The same method was later employed in generating and screening ~110,000 yeast double mutants to identify sensitivities to genotoxins. The yeast fitness was determined by comparing colony size in untreated conditions and in the presence of three chemical agents that induce DNA damage: DNA alkylating agent methyl methanesulfonate, topoisomerase I inhibitor camptothecin and DNA intercalating agent zeocin.<sup>146</sup> One potentially interesting advantage of the SGA system is that compelling double mutants could be back-crossed with further deletion mutants to generate triple mutants, which can be screened to investigate whether the lethal or drug-sensitive phenotype can be suppressed by a third mutation. By identifying potential suppressors of the synthetic lethal or synthetic sensitive phenotype it may be possible to anticipate and thereby prevent resistant tumours from developing.<sup>29</sup>

The alternative approach to combinatorial screening utilises RNAi, small molecules or genetic alterations to develop a sensitised cell background, which is submitted to large-scale RNAi screening. Any differences in response of isogenic cell lines to gene silencing are indicative of a synthetic genetic relationship, be that negative (synthetic lethal) or positive (suppression). For example, large-scale loss-of-function screens have been used to identify genes required by KRAS mutant colorectal cancer cells but not wild type cells lacking the oncogene. KRAS-transformed cancer cells showed synthetic lethality with proteasome and topoisomerase components<sup>147</sup>, and



CDK1 (cyclin-dependent kinase 1).<sup>148</sup> It is hoped that the datasets established from these screens will reveal new drug targets for use in deletion-guided cancer therapies.

Although synthetic lethality screening is now widely acknowledged as a valuable tool in identifying cancer vulnerabilities, a limited number of synthetic interactions have progressed from basic research. It is likely that the screening assays typically employed are limited by sensitivity and selectivity. In 2016, Hopkins *et al.* presented a novel method for synthetic lethality screening.<sup>3</sup> They explain that the majority of screens to date utilise pooled RNAi reagents, containing tens of thousands of RNA sequences. As such, the individual transfection frequency is low, resulting in poor sensitivity, inherent variability and poor reproducibility, as discussed previously, which could mean more modest synthetic interactions are overlooked. Accordingly they designed a screening system based on shRNA and siRNA double knockdown, in which isogenic cell lines expressing experimental or control shRNA with individual fluorescent tags were co-cultured and siRNA arrayed screening was used to diagnose synthetic lethal interactors, Figure 1.11. The system is intended for general use, but in this instance was developed to identify genes which are synthetic lethal with tumour suppressor BAF180 (BRG1-associated factor 180). Mutations in BAF180 are frequently observed in a variety of human cancers and in particular, it is a major ccRCC (clear cell Renal Cell Carcinoma) cancer gene, exhibiting truncating mutations in 41% of primary samples.<sup>1</sup> As such, it has been highlighted as a promising opportunity to target ccRCC using a synthetic lethal therapeutic approach.<sup>2</sup> A synthetic lethal interaction was discovered between BAF180 and histone acetyltransferase TIP60 (Tat-interactive protein, 60 kDa). Further data derived using this novel screening method is discussed in Chapter 4.



**Figure 1.11. Schematic of the Syngeneic Co-culturing System.** Two syngeneic cell lines, expressing shRNA for the gene of interest or control shRNA and labelled with GFP or mCherry respectively, were mixed together for use in an siRNA screen. Imaging and analysis of the GFP:mCherry ratio upon experimental siRNA treatment versus non-targeting control enabled validation of synthetic lethal pairs. Figure adapted from Hopkins *et al.*, 2016.<sup>3</sup>

The RNAi based technologies discussed have widely dominated research applications involving experimental modulation of gene expression over the past decade. However, recently a new gene editing technology, clustered regularly interspaced short palindromic repeat (CRISPR)-CRISPR-associated protein 9 (Cas9), has generated excitement. While RNAi regulates gene expression at the post-transcriptional level, the CRISPR-Cas9 system is used to introduce precision insertions and deletions directly to the genome. Due to the potential capabilities of the CRISPR-Cas9 system it has rapidly been accepted by the scientific community for use in a variety of genetic applications. It is likely that this technology will soon occupy many of the roles currently served by RNAi.<sup>149</sup>

## 1.4 Target Identification

Target identification is arguably the most crucial step in modern drug discovery. As discussed previously, drugs typically fail in clinic either because they do not work or they are not safe, and this can often be attributed to poor target selection.<sup>8</sup>

### 1.4.1 Exploratory Research

Drug discovery projects are typically initiated to address unmet clinical need. Suitable targets can be identified using a systems approach, a molecular approach, or often these in combination. The systems approach utilises whole organism studies in target discovery. Historically it was the main strategy for target identification and it remains so in diseases in which the relevant phenotype is only detectable at the organismal level. The molecular approach relies on establishing understanding of the cellular mechanisms underlying disease phenotypes to identify new targets. The approach therefore employs disease-relevant cells.<sup>150</sup> Outstanding advancements in molecular biology have caused a significant shift towards molecular approaches over the past two decades.<sup>151</sup> ‘Target’ is a broad term, encompassing an assortment of biological moieties, including molecular entities, proteins, genes or RNA, or disease biomarkers or biological pathways. A good target, as well as satisfying the unmet clinical need, must be efficacious, safe and druggable, i.e. accessible to a conceivable drug molecule.<sup>8</sup> It is well established that some target classes are more amenable to small molecule drug discovery, for example enzymes and G-protein-coupled receptors (GPCRs)<sup>152</sup>, but there is a necessity to explore more challenging targets in pursuit of improved therapeutics.

In initial research, data is generated to develop hypotheses that inhibition or activation of a particular protein or pathway will afford a desirable therapeutic effect. For instance, in the example above, TIP60 was identified as synthetic lethal with BAF180, which is frequently mutated in ccRCC cancer cells.<sup>3</sup> This suggests that BAF180-deficient tumours should be

selectively sensitive to TIP60-targeted therapeutics. The aforementioned rapid advances in biomolecular research combined with information from genome sequencing, gene expression, molecular interaction experiments, association studies and phenotypic screening have resulted in an explosion of such available data. As such, the challenge of prioritising targets for further experimentation is not trivial. For the best results, bioinformatics approaches should be adopted in data mining the available biomedical data, to identify, prioritise and ultimately select suitable targets.<sup>153</sup>

During the course of this project, attention has moved between a number of different biological targets, which form the basis for the thesis chapters. Consequently, the specific therapeutic relevance, regarding the efficacy and safety of each target is discussed in the chapter introductions. Common to all however is the way in which proteins are assessed for viability as druggable candidate targets.

#### 1.4.2 *Target Viability Data*

Advances in crystallographic methods, computational power, molecular biology and recombinant protein expression systems have resulted in much improved accessibility of three-dimensional (3D) structural information on a wide variety of proteins. This structural information can streamline many aspects of drug discovery such as target identification, the discovery of novel molecular scaffolds to form the basis for potential drugs, and lead optimisation. Accordingly, structure-guided drug design has been used to good effect in multiple therapeutic areas.<sup>9</sup> Most interesting in this instance is the use of structural information early in drug discovery to evaluate the viability of a protein as a drug target. Given the aim in drug discovery to design bioactive molecules, which selectively target a protein implicated in a diseased state without causing adverse effects, there is a requirement for the development of molecules with a controlled interaction profile against a variety of proteins. During target identification, validation and early lead discovery the focus is typically on the main target, but this shifts to include related and similar proteins as the drug discovery program progresses to lead optimisation.<sup>154</sup> Structure-based methods can be used to identify binding pockets in the target protein which might have suitable properties to accommodate a small molecule inhibitor and later to establish means to adequately discriminate between any related proteins to avoid undesirable off-target binding.<sup>9</sup>

Data accessibility is vitally important in the success of drug discovery projects. With such tremendous numbers of organic molecules, biological sequences and three-dimensional structures reported each year in scientific literature, there is a crucial need for it to be collected

and organised into databases.<sup>155</sup> Usefully, thousands of biological databases are detailed in the Nucleic Acids Research (NAR) online Molecular Biology Database Collection.<sup>156</sup> Some of these databases have been found to be of particular value in this work in druggability analysis. The Universal Protein Resource (UniProt) protein sequence database<sup>157</sup> is a comprehensive and freely accessible database of protein sequence and functional information. The UniProt Knowledgebase (UniProtKB) is comprised of the manually annotated UniProtKB/Swiss-Prot section and the automatically annotated UniProtKB/TrEMBL. The Protein Data Bank (PDB; [www.rcsb.org](http://www.rcsb.org))<sup>158</sup> is the single worldwide archive of three-dimensional structural data of biological macromolecules. ChEMBL<sup>159</sup> is a manually curated chemical database of drug-like molecules, which contains compound bioactivity data against drug targets. canSAR<sup>160</sup> is an integrated cancer research and drug discovery resource developed to improve data accessibility by gathering together publicly available biological annotation, information from screening, expression and amplification experiments and 3D structural data. Finally, STRING v9.1 (<http://string-db.org>)<sup>161</sup> is a database of known and predicted protein interactions, including physical and functional associations.

#### 1.4.3 *Structure-Based Druggability Assessment*

With growing numbers of available X-ray crystal structures and enhanced computational power, structure-based druggability assessment is often employed in target identification. Early assessment of druggability at target identification focuses efforts on targets more susceptible to therapeutic intervention by exposing potential target liabilities. This reduces drug discovery attrition, preventing unnecessary investment of time or money.<sup>9,162</sup> Target druggability analysis assesses the likelihood that a target is amenable to functional regulation by interaction with a drug-like molecule.<sup>162</sup> There are many computational methods that enable rapid and robust evaluation of druggability.

All drugs must not only interact with their molecular target but also be able to reach their site of action. Oral drugs, which must be absorbed from the gastrointestinal tract, typically therefore share specific physicochemical properties, as described by Lipinski's Rule of 5 (Ro5). Molecules should contain fewer than 5 H-bond donors, between 3 and 10 H-bond acceptors, have molecular weight less than 500 and have CLogP less than 5.<sup>163</sup> Most existing drugs engage their target proteins at predefined ligand binding sites, suggesting that compounds similar to the endogenous ligand should possess biological activity.<sup>152</sup> Furthermore, for the vast majority of these targets the substrate, product or allosteric effector is Ro5 compliant. However, this may be due to ease of design and the absence of a Ro5 ligand does not guarantee that a target is not druggable.<sup>162</sup> This is fortunate as most endogenous ligands are not drug-like, with less than a

third of the proteins annotated with ligands in the Human Metabolome Database<sup>164</sup> associated with a Ro5 compliant ligand.<sup>162</sup>

Methods of structure-based assessment of target druggability are typically assembled of three elements; a mechanism to detect potential binding sites, a mechanism to evaluate these sites based on their physicochemical properties, and a set of reference targets used to validate and refine the assessment.<sup>162</sup> Pocket-finding algorithms, used in predicting the ligand-binding sites, can rely purely on geometry or include physicochemical considerations. Fpocket, which is used in this work, utilises a geometry-only method to predict binding sites then goes on to score these based on the physicochemical properties of the surrounding atoms.<sup>165</sup>

Fpocket is an open source package which employs Voronoi tessellation and alpha spheres in pocket detection.<sup>165</sup> Voronoi tessellation is a method of dividing space with regard to a set of predefined points. In this instance qvoronoi from the qhull package<sup>166</sup> is utilised to produce a set of Voronoi vertices and the radii of alpha spheres centred at these positions is measured. An alpha sphere is a sphere which contacts four atoms at its boundary and contains no internal atoms. The four atoms will be, by definition, equidistant to the centre of the alpha sphere, the sphere radius. The radius of an alpha sphere is therefore dependent on the local curvature of the protein, as described by the four atoms. Very small spheres exist at the interior of proteins and very large spheres at the exterior, while clefts or cavities result in alpha spheres of intermediate radii, allowing them to be detected. By identifying clusters of such alpha spheres of intermediate radii, fpocket is able to discern structural pockets. Following this the spheres are categorised depending on the four surrounding atoms to enable filtering of the clusters with respect to their physicochemical properties, e.g. by hydrophobicity.<sup>165</sup>

It is good practice to validate structure-based target druggability methods against a set of reference targets of known tractability. Huang and Schroeder compiled a standard test set of 48 protein targets for which bound and apo crystal structures were available and utilised these to validate pocket-finding methods.<sup>167</sup> The average success rate of correct prediction of the true ligand binding site as the top scoring pocket was 60% for the apo structures and 67% for the bound structures. Fpocket in comparison achieved 69% for the apo structures and 83% for the bound structures.<sup>165</sup> Fpocket is also one of the few target druggability methods that is validated against a set of reference targets with known degrees of druggability. It can therefore be used to quantitatively assess druggability. This approach was spearheaded by Hajduk *et al.* who designed a simple model to assign the pockets a druggability score.<sup>168</sup> They included terms for polar and apolar surface area, surface complexity and pocket dimensions. The scores were tailored to conform with hit rates from nuclear magnetic resonance (NMR)-based fragment

screening, as an index of binding site druggability. The resulting algorithm was validated against a test set of 23 proteins with 57 pockets, which were not used in the training set, and correctly classified 94%. This reliable algorithm enables a quantitative assessment of the capacity of a given pocket to bind small compounds with high affinity and specificity. Cheng *et al.* developed a different but complementary approach two years later.<sup>169</sup> They utilised a biophysical binding-free-energy model, predominantly dependent on the curvature and hydrophobic surface area of the binding pocket, to devise the maximal affinity predicted for a passively absorbed oral drug ( $\text{MAP}_{\text{POD}}$ ) for the target. A test set of 27 pharmaceutical targets was compiled with 17 classified as druggable, 6 as difficult and 4 as undruggable. They found that a 100 nM threshold for  $\text{MAP}_{\text{POD}}$  score clearly separated those that are druggable from those which are difficult or undruggable, affording a mechanism to quantitatively assess druggability in other targets. Their manually curated dataset is now considered the benchmark for developing and validating new algorithms.<sup>162</sup> The validation set utilised in training fpocket is notable as it was compiled via an open collaborative platform (<http://fpocket.sourceforge.net/dcd>) and is the largest publicly available.<sup>170</sup> It combines targets from the two aforementioned studies with others manually annotated from the PDB.<sup>158</sup> The fpocket logistic model was trained from this dataset employing local hydrophobic density, hydrophobicity and normalised polarity as pocket descriptors.<sup>162,165</sup>

In the preceding studies, druggability is assessed solely on the properties of the protein. Alternatively, or additionally, the energetics of protein-ligand binding can be calculated via docking or molecular simulation. Molecular docking facilitates prediction of the binding modes and associated affinities of ligands at putative binding sites on the protein. Huang and Jacobsen virtually screened ~11,000 diverse fragments against 152 binding sites using docking to calculate computational hit rates for each site, indicative of druggability.<sup>171</sup> They showed that these hit rates correlated with those previously published from NMR-based screening.<sup>168</sup> However, this evolving technology still faces significant challenges that affect the accuracy of results, particularly in sampling approaches and scoring functions.<sup>172</sup> Due to the flexibility of the proteins and ligands, there are too many potential binding conformations to sample every one and although flexible docking is prevalent, it is computationally expensive. The scoring function is a prediction of ligand binding affinity. Current scoring algorithms underestimate the contributions of entropy and structural water and surrounding ions.<sup>173</sup> Seco *et al.* apply a decidedly different technique utilising molecular dynamics (MD) simulations of the interaction of the protein with isopropyl alcohol.<sup>174</sup> Analysis of the simulations enables deduction of the interaction free energies between the protein and probe molecule, which are used to detect binding sites and predict the maximal affinity of drug-like molecules at these positions.

MD simulations are now increasingly often being used in druggability assessment. Proteins are dynamic but a single X-ray crystal structure typically embodies only one of many possible conformations, although the B-factor provides a measure of the displacement of atoms from their mean position. The alternative conformations are not well evaluated using standard structure-based target druggability assessment. Structural variability can be scrutinised either by assessing multiple crystal structures of one target or by evaluating MD trajectories. Somewhat surprisingly, the druggabilities calculated from different crystal structures of the same protein can vary a great deal. Often this might be attributable to structural perturbation, such as mutant or missing residues, but in many cases it is the result of protein flexibility. A computationally expensive but thorough approach involves scoring the druggability of a series of structures along MD trajectories.<sup>162</sup> This highlights any plausible conformational changes that could occur under conditions of temperature, pressure, solvation and pH, simulating those *in vivo*, which may affect the druggability of the target.<sup>173</sup>

#### 1.4.4 *Modelling Three-Dimensional Protein Structure*

As discussed, thanks to the advances in crystallographic methods, there is now a vast amount of protein structural data available. However, the number of known protein sequences remains significantly higher than the number of solved structures. This is known as the sequence-structure gap.<sup>154</sup> Fortunately, for many of the proteins for which there are no published 3D structures, it might be possible to use protein structure prediction to enable structural analysis. Indeed, modelling techniques have now matured to a point of routine use in complementing experimental techniques.<sup>175</sup>

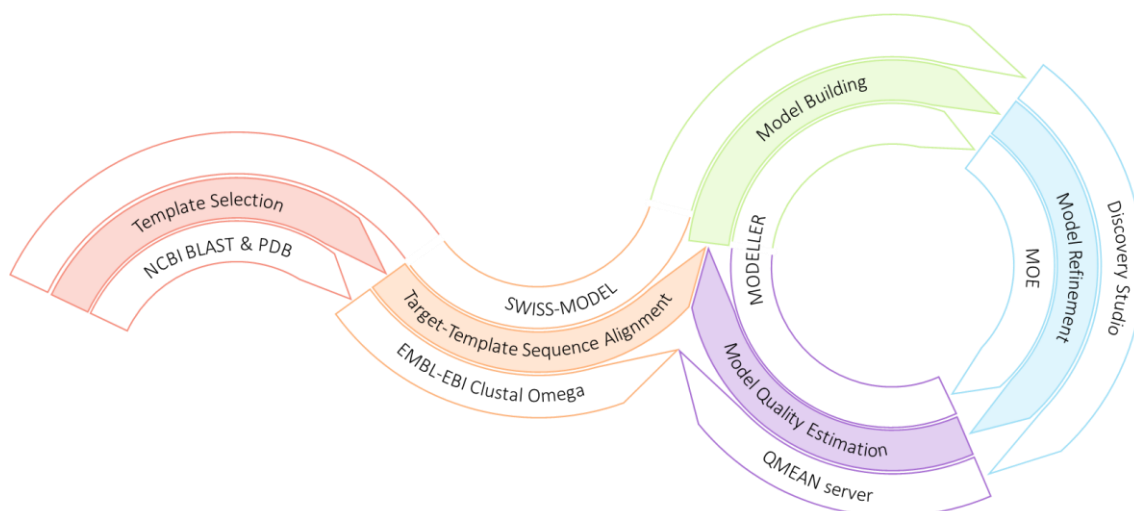
Computational methods for protein structure modelling are widely used in the pharmaceutical industry and a great deal of time and effort has been devoted to expanding the scope and improving the accuracy of the models. Current methods can be categorised as one or a combination of three approaches.<sup>176</sup> Firstly, homology or ‘comparative’ modelling; in which a model of the target protein of interest is generated based on protein sequence alignment with a homologous protein for which an experimental structure is available as a template. This technique is derived from the work of Šali and Blundell.<sup>177</sup> It relies on the observation that evolutionarily related sequences typically adopt similar 3D structures as they retain folds characterised by core structures that are robust against sequence modifications. It is therefore most effective when the structure of a closely related protein family member is available.<sup>154</sup> Secondly, where no structures of proteins with significant sequence similarity are available, fold recognition or ‘threading’ methods can be utilised. In threading, the target protein sequence is systematically aligned to a library of proteins of known structure and the fit is assessed by

energetics. The best match, producing the most reliable model, is identified as that with the lowest quasi-energy score, which represents the structural similarity of the target and template proteins. This approach is reliant on the hypothesis that because there are more known proteins than folds; the folds of a protein with unknown structure are likely to resemble known folds.<sup>176</sup> Thirdly, when neither comparative modelling nor threading can be utilised due to a lack of available templates, *de novo* methods can be employed. *De novo* methods are used to predict the protein structure directly from the primary sequence using the physical principles of protein folding. Information from determined structures may be incorporated but without assumption of any evolutionary relationships. Whilst versatile, this process is hugely computationally demanding and therefore successes tend to be limited to predicting folding of short peptides.<sup>176</sup> Of these approaches, homology modelling is considered the most accurate and is therefore the most commonly employed in drug discovery research.<sup>178</sup>

Template-based protein structure prediction is founded on two assumptions: that similar protein sequences will adopt similar folds, and that individual regions of a protein will exhibit the same folds already observed in the PDB. The process of homology modelling involves several phases: template selection, target-template sequence alignment, model building, model refinement and model quality estimation, Figure 1.12.<sup>154</sup> These stages are amenable to implementation of automated pipeline workflows, in which the user inputs the target protein sequence and the pipeline outputs a predicted structure.<sup>176</sup> For example, the SWISS-MODEL server (<http://swissmodel.expasy.org>)<sup>179</sup> enables automated comparative modelling of 3D structures.

In template selection, the experimental structure most appropriate for modelling is identified. Generally this is the 3D structure of the most closely related protein available, however there are additional criteria that might be considered. Single experimental structures and, by virtue of this, models only represent one conformation of the large range exhibited by proteins, which, as discussed, are intrinsically highly dynamic. Given that proteins can undergo a substantial rearrangement in order to accommodate a ligand, in some cases it can therefore be favourable to use an experimental structure depicting the protein in a ligand-bound state.<sup>154</sup> In addition, the quality of the template structure should be taken into account. Crystal structures with poor resolution or which are missing critical residues or loops may not provide adequate structural information to build a reliable model. In practice, programs such as NCBI BLAST (Basic Local Alignment Search Tool)<sup>180</sup> are used to search the PDB for suitable templates using the protein sequence of the target, and the most appropriate is selected.





**Figure 1.12. Homology Modelling Pipeline.** A typical homology modelling workflow of template selection, target-template sequence alignment, model building, model refinement and model quality estimation, annotated with suggested software for each step.

Target-template sequence alignment is performed by programs such as EMBL-EBI Clustal Omega.<sup>181,182</sup> The sequence identity of the target-template alignment is good indication of the quality of the resulting model. Models where the target-template sequence identity is above 50% are regarded as highly accurate and can be employed in drug discovery research. The protein core is typically modelled with high accuracy due to good evolutionary conservation in this domain and any flaws will likely only be observed in the packing of side chains or loop regions. Models based on less than 30% sequence identity are considered low accuracy models. Ambiguous alignment becomes a severe problem and it is possible that an entirely incorrect fold can be predicted.<sup>176</sup> In attempt to prevent this, more sensitive methods can be employed in homologue detection, such as DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST).<sup>183</sup> However, threading methods may provide better results in these instances.

Model building can be accomplished by two approaches. The first, rigid fragment assembly, was first implemented in 1987 by the Blundell lab.<sup>184</sup> The core of the model is constructed first from the best structurally conserved regions of one or more templates. Any inserts or deletions in the target-template alignment, such as loops, are then incorporated as fragments, each individually modelled on a template of close resemblance.<sup>185</sup> The SWISS-MODEL server employs a rigid fragment assembly modelling procedure.<sup>179</sup> The second is satisfaction of spatial restraints. Here spatial restraints are derived from a range of sources, including the target-template alignment, other known protein structures and molecular mechanics force fields. The target protein is then folded into the conformation which best satisfies these restraints. The most widely adopted approach that uses satisfaction of spatial restraints, which is also the industry standard in homology modelling, is MODELLER (<https://salilab.org/modeller/>).<sup>177</sup>

Following generation of primary protein models, refinement is required to optimise geometry and stereochemistry and remove any unfavourable contacts. This typically involves energy minimisation utilising a molecular mechanics force field, which may be followed by molecular dynamics to improve side chain contacts and rotamer states, and Monte Carlo sampling to improve accuracy of backbone conformations and core side chains.<sup>176</sup> Model building software often incorporates model refinement and evaluation capabilities. For further optimisation, drug discovery software such as Discovery Studio<sup>186</sup> and MOE (Molecular Operating Environment)<sup>187</sup> offer automated preparation functions to prepare and minimise models.

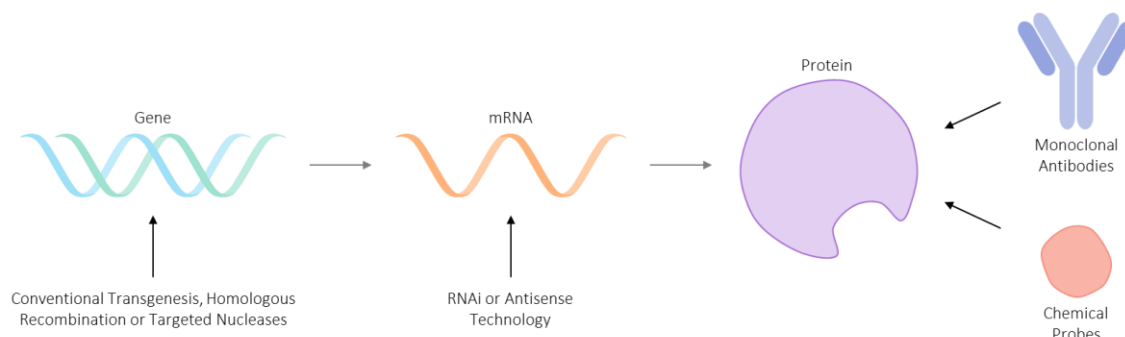
Model quality estimation is important as models can be produced with significant inaccuracies.<sup>154</sup> The geometrical accuracy and completeness of the model is evaluated and it is determined whether the proposed structure is energetically reasonable. Models can be assessed as to whether they possess the correct folds, which can aid in detecting errors in template selection, fold recognition and target-template alignment.<sup>188</sup> A scoring system, such as the DOPE (Discrete Optimised Protein Energy) score incorporated into MODELLER<sup>189</sup>, can be used to identify the best of a series of models. The quality of model required is highly dependent on its intended use, and in some instances, even if it is low resolution, 'any level of physical characterisation of a protein, as opposed to its absence, is valuable'.<sup>190</sup> Lower accuracy models can be sufficient in designing mutagenesis experiments or in preliminary target validation, whereas greater accuracy is required for structure-based virtual screening applications.<sup>191</sup>

Homology modelling capabilities are constantly improving but there are a few challenges that remain to be overcome. The accuracy of template-based modelling is limited by the availability of appropriate template structures and even at high sequence identity, although overall protein folds are well conserved, substrate specificity and mechanisms of catalysis vary greatly, indicating structural divergence. Approaches that specifically scrutinise and refine local structure could be used to complement homology-based modelling in these instances. There are also prevailing difficulties in refining models away from the template and toward the target structure, especially at low target-template sequence identity where significant rearrangement can be necessary.<sup>176</sup>

## 1.5 Target Validation

Once a suitable target has been identified, it should be investigated thoroughly to establish its clinical relevance before investment in costly drug discovery programmes. Such target validation should reduce attrition rates in the clinic and thereby help to improve R&D productivity. A large range of validation techniques are available, from *in vitro* tools, to mouse models, to patient

screening, which all have unique strengths and limitations. There is a strong temptation for groups to use only the approach most compatible with their technical expertise. However, while individual results are persuasive, a multi-validation approach is necessary to establish confidence in the target, Figure 1.13.<sup>8</sup>



**Figure 1.13. Overview of the Techniques Utilised in Target Validation.** mRNA expression can be modulated either directly by RNAi or antisense technology, or genetically in transgenic animals or xenograft models, by conventional transgenesis, HR or targeted nucleases. Target protein function can be modulated by monoclonal antibodies or chemical probes. Figure adapted from Lindsay, 2003.<sup>150</sup>

#### 1.5.1. Genetic Modulation

Transgenic and knockout animals are a useful validation tool, which enable elucidation of the functional consequence of gene manipulation.<sup>8</sup> Additionally, it is important to include an *in vivo* validation step in the early stage of drug discovery to verify that modulating the target retains an antitumourigenic effect in this setting, and to reduce the risk of missing targets that cannot be effectively assessed *in vitro*. Mouse models are the primary model of choice in cancer research. The models can be tailored for use in target validation, to assess whether a target is therapeutically relevant and loss of its function will be beneficial, and whether suppressing the target can inhibit the disease without causing systemic toxicity. The target is modulated by genetic means, either by directly manipulating the mouse germ line to produce genetically engineered mouse models (GEMMs), or by genetically manipulating human cancer cell lines, which are then transplanted into immunodeficient mice, termed xenograft models.<sup>192</sup>

There are two techniques used to produce GEMMs. The first is conventional transgenesis, in which plasmid DNA or bacterial artificial chromosome (BAC) DNA is injected into the pronucleus of fertilised oocytes, and these microinjected embryos are implanted into the oviducts of foster mother mice. Some of the resulting newborns will exhibit exogenous DNA randomly integrated into their genomes. These transgenic founder mice can be bred with wild-type mice to establish a selection of transgenic mouse lines. This method is typically used to insert new genetic information into the genome, to overexpress or knock down genes.<sup>192</sup> The first transgenic mouse lines were recorded by Constantini and Lacy in 1981.<sup>193</sup> The second is modification of the genomes of embryonic stem (ES) cells by homologous recombination. Targeted genome

modification is achieved with carefully designed DNA constructs. Cells in which the exogenous DNA sequence and genomic DNA have been successfully recombined are introduced into a developing mouse embryo, which is then implanted into the foster mother. The resulting newborn will be a chimeric mouse. The GEMM is produced when the genetic modification is transmitted to the offspring. This approach is typically used to manipulate single genes by knockout, knockin or point mutation.<sup>192</sup> The first targeted knockout mouse was achieved in 1989.<sup>194</sup> In gene knockins, an endogenous protein is replaced with a non-enzymatically functional protein. These animals may exhibit a different phenotype to the knockouts, particularly when the protein has structural functions as well as enzymatic.<sup>195</sup> Theoretically, gene knockin should better mimic the effect of drug treatment, as the protein is still present but not functional.<sup>8</sup>

The HR technology in particular has revolutionised mouse modelling and as testament to its huge scientific impact it was awarded the Nobel Prize for Medicine in 2007.<sup>192</sup> Mouse genomes can be precisely altered, for example to generate ‘oncomice’ in which to investigate gene alterations associated with cancer progression. Crossbreeding of these oncomouse GEMMs with appropriate tool GEMMs enables interrogation of proposed oncology targets *in vivo*, with the added genetic modification mimicking the therapeutic effect of a target inhibitor. For example, GEMMs were used in ascertaining that Cdk4 is synthetic lethal with KRAS<sup>196</sup>, and that Cdk inhibition is therapeutically beneficial in KRAS-driven non-small cell lung cancer.<sup>197</sup> This strategy also exposes any possible mechanism-based toxicity *in vivo*. Tool GEMMs are therefore profoundly enabling in biomedical discovery. Accordingly, the Knockout Mouse Project (KOMP) was initiated in 2004 as an international effort to produce and phenotype knockouts for all mouse genes and deliver them to the public domain.<sup>198</sup> This effort is ongoing and in 2016 the National Institutes of Health (NIH) awarded the University of California, Davis, over \$29 million for the next five years of related research.<sup>199</sup>

As research progressed, there was increasing demand for tissue restricted or inducible knockout mice, termed conditional GEMMs. Sometimes constitutive loss of function of a gene lead to embryonic lethality, preventing characterisation of the adult mice.<sup>138</sup> In addition, chronic absence of gene function could induce genetic adaptation by compensatory mechanisms and developmental phenotypes.<sup>8</sup> Temporally or spatially restricted gene expression is achieved by crossbreeding two GEMMs to merge the genetic modifications into one mouse. One allele consists of a silent mutation, which is stimulated only when an inducer is expressed or activated. For example, one widely used system is the Cre-LoxP system derived from bacteriophage P1.<sup>200</sup> The Cre recombinase specifically recognises 34 base pair loxP (locus of X-

over P1) sites. Mice are generated to harbour a gene of interest (GOI) flanked by loxP. These so called 'floxed' DNA sequences will be deleted wherever Cre is present. For spatial control, either the mice can be crossbred with transgenic mice expressing Cre in a cell or tissue specific manner, or Cre can be sporadically expressed at sites of interest via a viral vector. For temporal control, they can be crossed with mice expressing CreERT2 fusion protein, in which Cre is fused to an oestrogen receptor variant (ERT2). CreERT2 is ligand-dependent and inactive in the absence of ligand 4-hydroxytamoxifen, which affords inducible control of Cre activity and therefore target gene function. The ligand is delivered when desired, topically, systemically or via drinking water.<sup>192</sup>

Although GEMMs are extremely useful in enabling biological studies and target identification and validation, they are expensive and time-consuming to produce. As such, attention has been drawn to the potential uses of non-germ-line GEMMs, human xenograft mouse models, generated by transplanting human cancer cells that have undergone *ex vivo* genetic engineering, into immunodeficient mice. These have the advantage over GEMMs that they are of human origin and therefore reproduce the genetic complexity of human tumours. To interrogate and validate proposed oncology targets, the cancer cells are genetically modified in such a way as to mimic the effect of inhibiting the target, and the resulting effect is analysed *in vivo*.<sup>192</sup> Advancing technologies facilitate the introduction of appropriate targeted mutations into the cancer cell DNA. Targeted nucleases enable the rapid and economical manipulation of genomic sequences in a broad spectrum of cell types and organisms. Zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and CRISPR-Cas9 are designed to recognise specific sequences in the genome. The nucleases generate DSBs at these specific locations and stimulate DNA damage repair mechanisms, which can be harnessed to achieve precisely targeted genomic edits.<sup>201</sup> Genome editing with engineered nucleases was awarded Method of the Year in 2011.<sup>202</sup>

#### 1.5.2. mRNA Modulation

The modulation of target genes in cancer cells for xenograft models can also be achieved by knocking down the expression of the GOI by RNAi, using either siRNA or shRNA. Generally it is preferable to use shRNA due to the challenges in delivery, stability and transient knockdown associated with siRNA. shRNA is delivered by transfecting or transducing cells with plasmids or viral vectors respectively, and is designed such that it is stably integrated into the genome, expressed under either a constitutive or inducible promoter. The cells are transplanted into immunocompromised mice to enable evaluation of the effect of downregulating the target *in vivo*. Again, inducible expression offers a number of advantages, including the prevention of undesired effects on cell growth and the opportunity for time-dependent regulation.<sup>203</sup>

Proposed targets are validated by genetic modification of the cells to imitate therapeutic intervention. Targets such as Ran (Ras-related Nuclear protein)<sup>204</sup>, receptor tyrosine kinase HER3 (Human Epidermal Growth Factor Receptor 3)<sup>205</sup> and PLC $\gamma$ 1 (phospholipase C $\gamma$ 1)<sup>206</sup> have been validated by this approach.

RNAi-based gene knockdown has also evolved into an alternative approach to classic gene knockdown in mouse models. As in cultured cells, siRNA can be used to transiently induce gene silencing in somatic tissues and is frequently employed in experiments requiring only short term knockdown of gene function.<sup>207</sup> Local delivery of siRNAs is also proving valuable in therapeutics and there has been a resurgence in clinical trials employing RNAi technology.<sup>208</sup> Permanent gene silencing is achieved by integration of shRNAs into the mice genomes. Transgenesis of the shRNA occurs as discussed above, by pronucleus injection, viral embryo infection or targeted knockin by homologous recombination to ES cell genomes. The shRNA expression vectors are inserted by recombinase-mediated cassette exchange (RMCE) into the ubiquitously active *Rosa26* locus of the ES cells, and these are developed into GEMMs. This targeted insertion ensures ubiquitous and reproducible expression of the shRNA, which affords constitutive gene silencing by mRNA degradation. Spatially- or temporally-specific silencing can again be achieved utilising the Cre-LoxP system. A loxP-STOP cassette, which disrupts function, is inserted at a restriction site in the shRNA loop region. The dsRNA structure of the shRNA is formed from carefully designed single-stranded RNA, comprising of a 19-22 base pair sense siRNA sequence linked to the complementary anti-sense sequence by a loop region of 8 bases. After transcription the complementary RNA sequences hybridise to form the dsRNA which is exported from the nucleus and processed by DICER to produce siRNA, Figure 1.10.<sup>207</sup> Despite significant developments in RNAi technologies, development of GEMMs nonetheless remains a costly and time-consuming endeavour. *In vitro* screening and target validation using RNAi has also highlighted many promising drug targets, however, *in vivo* validation is desirable before investing in a full drug discovery program.

Similarly to RNAi, antisense technology can be used to silence genes by transcriptional regulation. RNA-like oligonucleotides are designed which are complementary to a region of the target mRNA.<sup>8</sup> The antisense oligonucleotides (ASOs) enter the cell, exploiting either natural processes or facilitators, and interact with the target mRNA. Formation of the ASO-mRNA duplex activates RNase H, which cleaves the mRNA, thereby impeding protein synthesis. Like siRNA, ASOs represent a promising opportunity in therapeutics and have been tested in more than 100 clinical trials.<sup>209</sup> Antisense technology can also be used in target validation. For example, antisense probes for the P2X<sub>3</sub> receptor induce reduced pain sensitivity in the Complete Freund's

Adjuvant (CFA) model, validating the role of the receptor in chronic inflammation.<sup>210</sup> Unlike gene knockout and shRNA approaches, antisense technology is reversible and after administration is discontinued, target function will be restored.<sup>8</sup> A number of challenges of ASOs remain to be addressed, including issues with delivery, targeting and off-target effects.<sup>209</sup>

#### 1.5.3. Protein Modulation

With the specific exception of gene knockin, all of the approaches described above result in the complete elimination of the target protein from the cell. This however may not be the best approach to validate that a target is amenable to regulation by small molecule drug compounds. A small molecule will likely inactivate the enzymatic function of a protein but not destroy it, or cause major structural disruption. Any structural functions of the protein, for example regarding complex formation, may therefore remain viable, despite inhibition of the enzymatic function.<sup>8</sup> In target validation, it is therefore prudent to include an effort in protein modulation.

Monoclonal antibodies (mAbs) are a useful tool in target validation. They interact with a large region of the target molecular surface, affording good selectivity and high affinity.<sup>8</sup> HER3 plays an essential role as a cell surface receptor for neuregulins. In contrast to other HER receptors HER3 lacks intrinsic kinase activity. As such, mAbs, which are able to prevent the receptor from interacting with ligands or other proteins, are the most suitable tool for HER3 target validation and therapeutics. Over a dozen anti-HER3 mAbs are in clinical development and clinical observations strongly validate HER3 as a cancer drug target.<sup>211</sup> While monoclonal antibodies provide high affinity and good discrimination between targets and therefore show little non-mechanistic toxicity, they are unable to cross cell membranes and as such the target class is restricted mainly to cell surface and secreted proteins.<sup>8</sup>

In many cases it is therefore necessary to design and prepare small bioactive molecules, which are able to functionally modulate the target protein.<sup>8</sup> It is important to note the clear distinction between molecules destined to become drug candidates and these chemical probes, which are used primarily to explore protein function and validate efficacious and druggable targets. It is often acceptable that probes exhibit less good pharmacokinetic properties, which offers a wider scope to incorporate diverse chemical groups<sup>212</sup>, but the probes must be highly selective. In fact, selectivity is often more important for chemical probes than it is for drug compounds, which may benefit from off-target activity via polypharmacology. It is necessary that the probes exhibit very few if any off-target effects, to enable clear and reliable interpretation of subsequent experimental data.<sup>213</sup> Unfortunately, there are many widely used chemical probes that do not comply with acceptable potency and selection criteria, and so afford unreliable biological data.

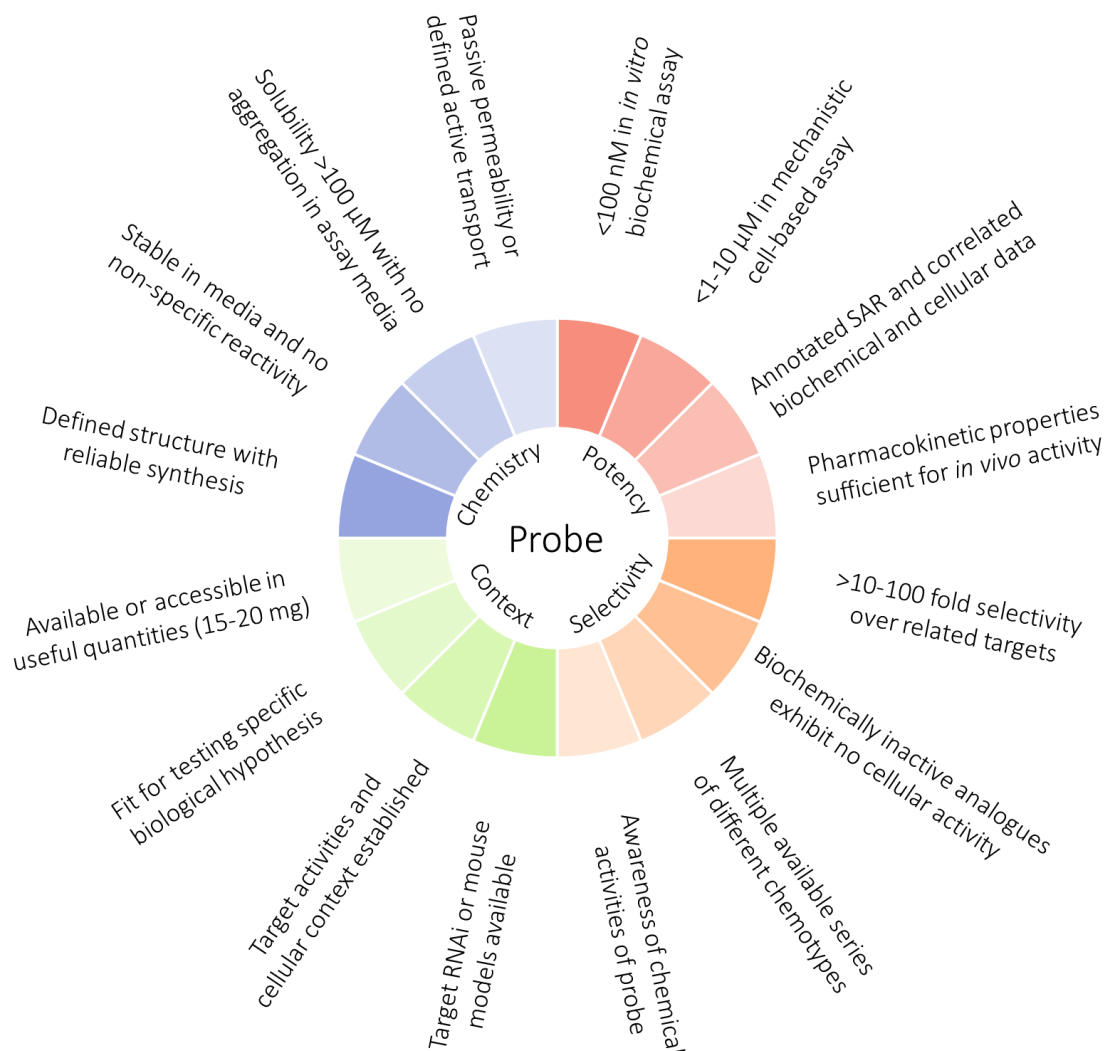
Bunnage *et al.* suggest four ‘pillars of target validation’, which increase the likelihood that drug programs succeed in the clinic.<sup>214</sup> They propose that by applying these principles to development and use of chemical probes, target perturbation can be more confidently linked to disease-relevant pharmacological regulation. Pillar one is sufficient exposure at the site of action, which requires that the probe exhibits good cellular permeability. As Lipinski’s Ro5 establishes criteria important in drug absorption<sup>163</sup>, there are also criteria acknowledged to improve cellular penetration, for example regarding lipophilicity.<sup>215</sup> By validating probe accessibility it is possible to rule out false negatives, where the lack of cellular response is the result of poor cellular penetration. Moreover, it may be necessary to determine the precise concentration of the probe inside cells, to accurately evaluate dose response and structure-activity relationship (SAR), or to ensure that the probe is maintained at a concentration within its selectivity window. Liquid chromatography-mass spectrometry (LC-MS) or microscopic imaging could be utilised in this endeavour.<sup>214</sup> Pillar two is proof of target engagement and selectivity. A thorough understanding of the performance of the probe in whole-cell and *in vivo* is vital, including both the true on- and off-targets. In-depth molecular mode-of-action studies and selectivity profiling help to increase confidence in the target and probe. An inactive analogue is highly useful as a negative control. Activity-based proteomics techniques enable quantification of target engagement. When a probe is developed, a bespoke chemical proteomics probe can be designed, founded on the structure of this tool compound. The proteomics probe should consist of the binding group (from the tool compound), which imparts selectivity, a protein-reactive warhead, to stabilise the protein-probe interaction, and a tag, which could be a reporter, such as a fluorophore, or an affinity label, such as biotin or an alkyne or azide for use in click chemistry. Application of the proteomics probe, followed by cell lysis and experimental interpretation, can confirm the cellular target or targets of the probe.<sup>214</sup> Pillar three is expression of functional pharmacology, i.e. the probe should affect the biochemical function of the target. An assay appropriate to the target must be selected with which to measure the affinity or IC<sub>50</sub>s of the candidate probes. Patient-derived cells can be particularly helpful as they often endure pathophysiological modifications that simpler model cell lines lack. Finally, pillar four is expression of a disease-relevant phenotype. As previously highlighted, often biomarkers can be disease-related but non-causal. Unearthing these early in preclinical research confers a substantial saving of time and money. Mammalian models are often prohibitively expensive and time-consuming, but more primitive organisms such as zebrafish can be employed in phenotypic screening.<sup>214</sup> Together these pillars help to determine the desirable characteristics of small molecule probes.



Workman and Collins define a series of 'fitness factors' as guidelines for consideration when assessing the quality of chemical probes, with an aim to avoid generation of poor quality and misleading data, while retaining the freedom for creative and innovative research.<sup>215</sup> These factors are separated into four categories, Figure 1.14. (1) Chemical properties. It is paramount that the probe is well-characterised and pure, and that reliable and reproducible preparative methods are available. It should be stable in relevant media and free from non-specific chemical reactivity. Solubility and membrane permeability are also of great importance if the probe is to be used in cell-, tissue- or whole organism-based research. It should be sufficiently soluble in aqueous media, not prone to aggregation in assay conditions and have proven passive membrane permeability or defined active transport mechanisms. (2) Biological potency. The potency of the probe should be appropriate for the intended use, with approximate guidelines of <100 nM in *in vitro* biochemical assays and 1-10  $\mu$ M in cell-based assays. Established SAR and good correlation between biochemical and cellular data improves confidence in the targeted mechanism of action of the probe. Finally, to be useful in *in vivo* studies, the probe must have suitable pharmacokinetic properties to achieve levels in target tissues that are relevant to the cellular potency. (3) Biological selectivity. A good chemical probe should exhibit >10-100 fold selectively over other related targets, and analogues with no measurable biochemical activity should also exhibit no activity in cells. Ideally there should be multiple probes with different chemotypes available, which exhibit similar activity, to improve confidence in data reliability, and any chemical activities of the probe should be well understood. (4) Context of use. In addition to the intrinsic properties of the compound, the use governs the suitability of a probe. It is favourable to have access to RNAi or mouse models of the target for complementary experiments. The cellular context of the target and any linked activities should be considered, as well as the fitness of the probe to test the specific biological hypothesis. Finally, the probe should be available or accessible in quantities that enable follow up studies (15-20 mg).<sup>215</sup>

Good quality chemical tools provide an excellent mechanism to test therapeutic hypotheses with a molecule close to the profile of a clinical candidate. Demonstration of a non-toxic therapeutic phenotype in cellular or *in vivo* models in response to the probe affords a significant boost in confidence and is fundamental in target validation.<sup>216</sup> Use of such chemical probes remains disadvantaged by the need to interact with conserved sites on the protein and the resulting high propensity for non-mechanistic toxicity.<sup>8</sup> However, careful target selection and thorough counter-screening efforts assist in relieving this difficulty, and inhibitor development is certainly a worthwhile venture. Chemical probes are remarkably powerful in unveiling and accessing novel biology and potential targets.<sup>214</sup> For example, the discovery of selective BET

(bromodomain and extra-terminal domain) inhibitors in 2010<sup>217,218</sup>, has resulted in an explosion of research in BET bromodomain biology, the therapeutic potential of BET inhibition and the role of other bromodomains.



**Figure 1.14. Criteria for Designing and Creating Effective Chemical Probes.** Fitness factors grouped into four areas: chemical properties, biological potency, selectivity, and context of use, which enable evaluation of the suitability of a chemical probe in exploratory biology. Figure adapted from Workman & Collins, 2010.<sup>215</sup>

## 1.6 Project Aim

The aim of this project was to design and characterise tool inhibitors of proteins involved in DNA damage response, primarily focusing on the structural maintenance of chromosomes (SMC) proteins and lysine acetyltransferase 2-A (KAT2A). It is hoped that these probes will emulate the success of others in enabling detailed mechanistic biological investigation, disease validation and ultimately initiation of translational drug discovery projects.

## 2 Materials and Methods

### 2.1 Bioinformatics

#### 2.1.1 Data Procurement

Information on proteins of interest was collected from online resources and databases; protein descriptions and domain information from canSAR 2.0<sup>160</sup>, protein association networks from STRING v9.1<sup>161</sup>, phylogenetic trees from the Structural Genomics Consortium website<sup>219</sup> and information regarding known small molecule inhibitors or binders from ChEMBL<sup>159</sup>.

The complete amino acid sequences of targets were retrieved, in FASTA format, from the UniProt protein sequence database.<sup>157</sup>

#### 2.1.2 Homology Searches

Utilising the amino acid sequences of targets, the NCBI BLAST<sup>180</sup> program was used to search for:

- Human homologues (Database: UniProtKB/Swiss-Prot, Organism: *Homo sapiens* (taxid: 9606))
- Orthologues (Database: UniProtKB/Swiss-Prot)
- Relevant crystal structures available in the PDB (Database: Protein Data Bank)

Commonly, BlastP is used to compare the protein query to a protein sequence database. In this work however, DELTA-BLAST was used, which searches a database of pre-constructed PSSMs (Position Specific Scoring Matrices) before searching the protein database, in order to yield better homology detection.<sup>183</sup> The complete FASTA sequences of selected hits were downloaded from the DELTA-BLAST search results.

#### 2.1.3 Sequence Alignments

EMBL-EBI Clustal Omega<sup>181,182</sup> was used to compile pairwise and multiple sequence alignments of protein sequences, utilising amino acid sequences downloaded from UniProt protein sequence database.<sup>157</sup>

#### 2.1.4 Conservation Scoring

The Scorecons server<sup>220</sup> was used to rate the conservation at each amino acid site in pairwise and multiple sequence alignments of:

- Human SMC proteins and Rad50
- KAT2A and KAT2B HAT domains
- GNAT and MYST HAT domains
- KAT2A and KAT2B bromodomains
- Type I and type II bromodomains
- All human bromodomains

Scorecons provides a measure of the site-specific conservation, using a sum-of-pairs scoring system to assign scores ranging from 0 to 1, which depend on the amino-acid frequency and relative stereochemical properties of substituted amino acids.<sup>220</sup>

## 2.2 Homology Modelling

### 2.2.1 Model Building

Modelling was performed using homology modelling programs, SWISS-MODEL<sup>179</sup> and MODELLER 9.12<sup>177</sup>, which rely on protein sequence alignment, and PHYRE 2.0 (Protein Homology/analogy Recognition Engine)<sup>221</sup> which uses a protein threading software.

SWISS-MODEL<sup>179</sup> was used to build one model, Table 2.1.

**Table 2.1. Models of Target *Homo Sapiens* Protein Domains Constructed Using SWISS-MODEL.<sup>179</sup>**

Target	Organism	Template		Resolution
		Protein & Domain	PDB ID	
SMC5/6 hinge	<i>Thermotoga maritima</i>	SMC hinge	1GXJ <sup>222</sup>	2.00 Å

MODELLER 9.12<sup>177</sup> was used to build seven models, Table 2.2.

**Table 2.2. Models of Target *Homo Sapiens* Proteins or Protein Domains Constructed Using MODELLER 9.12.<sup>177</sup>**

Target	Organism	Template		Resolution
		Protein & Domain	PDB ID	
SMC5/6 hinge	<i>Schizosaccharomyces pombe</i>	SMC5/6 hinge	5MG8 <sup>223</sup>	2.75 Å
SMC5/6 head	<i>Pyrococcus furiosus</i>	SMC ATPase	1XEX <sup>224</sup>	2.50 Å
NSE1	<i>Homo sapiens</i>	NSE1	3NWO A <sup>225</sup>	2.92 Å
NSE2	<i>Saccharomyces cerevisiae</i>	Mms21	3HTK C <sup>226</sup>	2.31 Å
NSE3	<i>Homo sapiens</i>	MAGEA4	2WAO <sup>227</sup>	2.30 Å
SMC1/3 hinge	<i>Mus musculus</i>	SMC1/3 hinge	2WDS <sup>228</sup>	2.70 Å
SMC1/3 head	<i>Pyrococcus furiosus</i>	SMC ATPase	1XEX <sup>224</sup>	2.50 Å

Models were constructed by the satisfaction of spatial restraints, using the ‘automodel’ class. Five similar models were generated using the python script command, additionally assessing the DOPE score<sup>189</sup> and GA341 score, and the ‘best’ selected as that with the lowest DOPE score, *i.e.* that with the lowest associated energy. Table 2.3 to Table 2.9 show the log file summaries for all the models built. The ‘best’ model, which was further analysed, is highlighted.

**Table 2.3. MODELLER 9.12<sup>177</sup> Log File Summary for SMC5-SMC6 Hinge Models.**

Filename	molpdf	DOPE Score	GA341 Score
HSMCAB_hinge.001.pdb	4286.1	-38403.9	1.000
HSMCAB_hinge.002.pdb	4023.1	-38437.8	1.000
HSMCAB_hinge.003.pdb	4131.2	-38639.1	1.000
HSMCAB_hinge.004.pdb	4217.1	-38916.5	1.000
HSMCAB_hinge.005.pdb	4008.4	-38514.0	1.000

**Table 2.4. MODELLER 9.12<sup>177</sup> Log File Summary for SMC5-SMC6 Head Models.**

Filename	molpdf	DOPE Score	GA341 Score
HSMC56_head.001.pdb	4875.5	-66034.6	0.921
HSMC56_head.002.pdb	4729.7	-66159.1	0.904
HSMC56_head.003.pdb	5003.4	-65486.4	0.850
HSMC56_head.004.pdb	5396.0	-64320.7	0.769
HSMC56_head.005.pdb	4935.7	-65919.4	0.917

**Table 2.5. MODELLER 9.12<sup>177</sup> Log File Summary for NSE1 Models.**

Filename	molpdf	DOPE Score	GA341 Score
HNSE1.B99990001.pdb	1584.1	-27186.2	1.000
HNSE1.B99990002.pdb	1488.4	-27065.1	1.000
HNSE1.B99990003.pdb	1408.9	-27149.6	1.000
HNSE1.B99990004.pdb	1430.3	-27134.1	1.000
HNSE1.B99990005.pdb	1529.6	-27031.7	1.000

**Table 2.6. MODELLER 9.12<sup>177</sup> Log File Summary for NSE2 Models.**

Filename	molpdf	DOPE Score	GA341 Score
HNSE2.B99990001.pdb	1360.3	-16265.0	0.448
HNSE2.B99990002.pdb	1470.3	-16566.1	0.845
HNSE2.B99990003.pdb	1371.4	-16093.6	0.522
HNSE2.B99990004.pdb	1315.5	-16638.1	0.462
HNSE2.B99990005.pdb	1401.7	-16193.6	0.485

**Table 2.7. MODELLER 9.12<sup>177</sup> Log File Summary for NSE3 Models.**

Filename	molpdf	DOPE Score	GA341 Score
HNSE3.B99990001.pdb	1596.2	-26809.6	1.000
HNSE3.B99990002.pdb	1503.4	-26783.0	1.000
HNSE3.B99990003.pdb	1569.1	-26738.9	1.000
HNSE3.B99990004.pdb	1565.5	-26527.8	1.000
HNSE3.B99990005.pdb	1589.6	-26791.4	1.000

**Table 2.8. MODELLER 9.12<sup>177</sup> Log File Summary for SMC1-SMC3 Hinge Models.**

Filename	molpdf	DOPE Score	GA341 Score
HSMCAB_hinge.B99990001.pdb	2628.6	-43777.5	1.000
HSMCAB_hinge.B99990002.pdb	2564.9	-44103.1	1.000
HSMCAB_hinge.B99990003.pdb	2606.3	-43488.2	1.000
HSMCAB_hinge.B99990004.pdb	2658.5	-43959.9	1.000
HSMCAB_hinge.B99990005.pdb	2583.3	-43485.1	1.000

**Table 2.9. MODELLER 9.12<sup>177</sup> Log File Summary for SMC1-SMC3 Head Models.**

Filename	molpdf	DOPE Score	GA341 Score
HSMC13_head.B99990001.pdb	4302.1	-77329.9	0.997
HSMC13_head.B99990002.pdb	4004.0	-77646.0	1.000
HSMC13_head.B99990003.pdb	4277.9	-78393.1	1.000
HSMC13_head.B99990004.pdb	4326.0	-77139.1	0.999
HSMC13_head.B99990005.pdb	4146.5	-77750.7	0.987

Phyre 2.0<sup>221</sup> was used to build one model, Table 2.10. A model confidence of 100% was quoted.

**Table 2.10. Models of Target *Homo Sapiens* Protein Domains Constructed Using Phyre 2.0.<sup>221</sup>**

Target	Organism	Template		Resolution
		Protein & Domain	PDB ID	
SMC5 head	<i>Deinococcus radiodurans</i>	RecN	4AD8 <sup>229</sup>	4.00 Å

### 2.2.2 Model Preparation and Minimisation

Preparation and minimisation of protein models was performed using Discovery Studio 4.0.<sup>186</sup> The Automatic Preparation function was used to prepare, clean and protonate the model, which was then minimised using the default parameters. Table 2.11 displays CHARMM (Chemistry at HARvard Macromolecular Mechanics) energies calculated for each model at minimisation.<sup>230,231</sup>

**Table 2.11. CHARMM Energies Calculated by Discovery Studio 4.0<sup>186</sup> at Model Minimisation.**

Model Filename	Modelling Software	CHARMM Energy
HSMCAB_hinge.004.pdb	MODELLER	-23880.0
HSMC56_head.001.pdb	MODELLER	-40835.0
HSMC56_head.002.pdb	MODELLER	-41080.4
HSMC56_head.003.pdb	MODELLER	-40831.4
HSMC56_head.004.pdb	MODELLER	-40564.7
HSMC56_head.005.pdb	MODELLER	-40858.1
HSMC5ab_head_4AD8.pdb	PHYRE 2.0	-18068.4
HNSE1.B99990001.pdb	MODELLER	-16412.2
HNSE2.B99990004.pdb	MODELLER	-15264.1
HNSE3.B99990001.pdb	MODELLER	-16664.6
HSMCAB_hinge.B99990002.pdb	MODELLER	-29243.0
HSMC13_head.B99990003.pdb	MODELLER	-46803.4

## 2.3 Structural Alignment and Visualisation

**Table 2.12. Structural Alignments Between Targets and Templates Prepared Using PyMOL.<sup>232</sup>**

Organism	Target		Organism	Template	
	Protein & Domain	PDB ID		Protein & Domain	PDB ID
<i>Homo sapiens</i>	SMC5 head	Model	<i>Pyrococcus furiosus</i>	SMC ATPase	1XEX <sup>224</sup>
<i>Homo sapiens</i>	NSE3	Model	<i>Homo sapiens</i>	NSE1-MAGEG1	3NWO B <sup>225</sup>
<i>Homo sapiens</i>	SMC1/3 head	Model	<i>Saccharomyces cerevisiae</i>	SMC1-SCC1	1W1W A <sup>233</sup>
<i>Rattus norvegicus</i>	RNF4	3NG2 B <sup>234</sup>	<i>Homo sapiens</i>	RNF4	2XEU <sup>235</sup>
<i>Homo sapiens</i>	ASF1A	2IO5 A <sup>236</sup>	<i>Homo sapiens</i>	ASF1A-HIRA	2I32 <sup>237</sup>
<i>Homo sapiens</i>	GCN5 BRD	3D7C <sup>238</sup>	<i>Saccharomyces cerevisiae</i>	GCN5 BRD-H4	1E6I <sup>239</sup>
<i>Homo sapiens</i>	PCAF BRD	5FE8 <sup>240</sup>	<i>Saccharomyces cerevisiae</i>	GCN5 BRD-H4	1E6I <sup>239</sup>

PyMOL 1.6<sup>232</sup> was used as standard to produce 3D representations of proteins and models and to prepare structural alignments utilising the *action\_align\_to\_molecule* functionality, Table 2.12.

## 2.4 Pocket Prediction and Scoring

Binding pockets were defined using both Discovery Studio 4.0<sup>186</sup> and Fpocket 2.0<sup>165</sup>, Table 2.13. In both programs the binding pockets were defined from receptor cavities using the default parameters. The structure-based target druggability prediction score coupled to Fpocket 2.0 was utilised in assessing sites.<sup>170</sup>

**Table 2.13. Targets of Pocket Prediction Using Discovery Studio<sup>186</sup> and Fpocket<sup>165</sup>.**

Organism	Target Protein & Domain	PDB ID	Prediction Software	
			Discovery Studio <sup>186</sup>	Fpocket <sup>165</sup>
<i>Homo sapiens</i>	SMC5/6 hinge	Model	Y	
<i>Homo sapiens</i>	SMC5/6 head	Model	Y	Y
<i>Pyrococcus furiosus</i>	SMC ATPase	1XEX <sup>224</sup>	Y	Y
<i>Homo sapiens</i>	SMC5 head	Model	Y	Y
<i>Homo sapiens</i>	NSE1	Model	Y	Y
<i>Homo sapiens</i>	NSE2	Model	Y	Y
<i>Homo sapiens</i>	NSE3	Model	Y	Y
<i>Homo sapiens</i>	SMC1/3 hinge	Model	Y	Y
<i>Homo sapiens</i>	SMC1/3 head	Model	Y	Y
<i>Saccharomyces cerevisiae</i>	SMC1-SCC1	1W1W A <sup>233</sup>		Y
<i>Rattus norvegicus</i>	RNF4	3NG2 B <sup>234</sup>		Y
<i>Homo sapiens</i>	ASF1A	2IO5 A <sup>236</sup>		Y
<i>Homo sapiens</i>	GCN5 HAT	1Z4R <sup>241</sup>		Y
<i>Homo sapiens</i>	GCN5 BRD	3D7C <sup>238</sup>		Y

## 2.5 Data Handling - KNIME

KNIME (Konstanz Information Miner)<sup>242</sup> is a modular environment that enables visual assembly and execution of data pipelines. Nodes are assembled into workflows for data processing.

### 2.5.1 Filter Results of HTS and Hit Confirmation Assay

A KNIME<sup>242</sup> workflow was created to filter the results of hit confirmation assay, PubChem BioAssay AID: 588347<sup>5</sup>, completed by the NIH Chemical Genomics Center (NCGC), to identify compounds from the set that were defined as 'active', had acceptable curve descriptions, discarding any with efficacy >150 calculated from a partial curve, and were available from preferred suppliers.

Of the 145 compounds screened, 82 were available from preferred suppliers, identified using workflow prepared by Dr Ben Wahab (Sussex Drug Discovery Centre, University of Sussex). This

filter output an SDF file detailing the structures of the available compounds. Meta Node ExtractInfoForSubset was used to reintroduce information from the assay summary, regarding the activity profile and curve descriptors. A row filter was utilised to select entries with PUBCHEM\_ACTIVITY\_OUTCOME 'Active', and Meta Node GoodCurveDescription was used to remove entries with unreasonable curve descriptors, where efficacy >150 was calculated from partial curves. Finally, Meta Node DiversityCluster was employed in attempt to cluster the compounds into series of similar structures and the data was output to an interactive table.

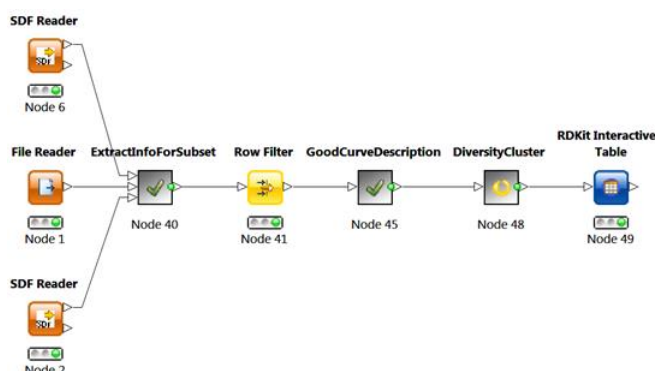


Figure 2.1. Overview of KNIME<sup>242</sup> workflow used to filter results of hit confirmation assay.

#### *ExtractInfoForSubset*

In Meta Node ExtractInfoForSubset, tautomers were generated of the compounds from both the original hit confirmation assay and the SDF file of those available for purchase. To the former an Orig\_number column was introduced using RowID to enable later identification of origin and to the latter, a constant value column to infer availability. The two datasets were then concatenated, grouped by Unique\_SMILES, filtered on availability from preferred suppliers, and grouped back by Orig\_number to give individual entries for each compound input. Finally, these individual entries were recombined with the non-structural information from the assay summary by concatenating the datasets and again grouping by Orig\_number and filtering on availability from preferred suppliers.

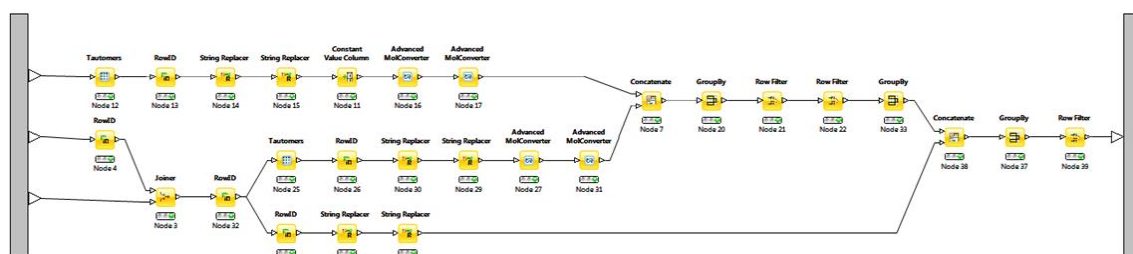


Figure 2.2. Outline of KNIME<sup>242</sup> ExtractInfoForSubset Meta Node.

#### *GoodCurveDescription*

In Meta Node GoodCurveDescription those entries with Curve\_Description 'Partial curve; high efficacy' were selected and filtered to retain only those with Efficacy lower than 150, before



recombining with entries with Curve\_Descriptions 'Complete curve; high efficacy', 'Complete curve; partial efficacy' and 'Partial curve; partial efficacy'.

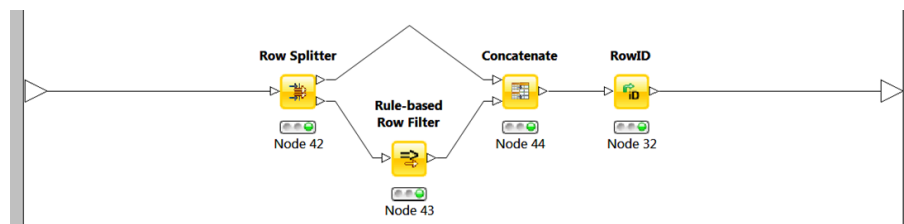


Figure 2.3. Outline of KNIME<sup>242</sup> GoodCurveDescription Meta Node.

### DiversityCluster

In Meta Node DiversityCluster, initially the RDKit Diversity Picker was used to select a diverse set from the data. A loop was initiated in which the fingerprints of the diverse molecules were compared to those of the complete dataset to calculate the Tanimoto similarity coefficient, the coefficients were sorted into descending order and entries with coefficients below 0.5 were removed. To collect only results which had been successfully clustered, the number of results of each iteration were counted and those with two or more associated results were selected. A second loop was then used to recollect the compounds that were identified in those iterations.

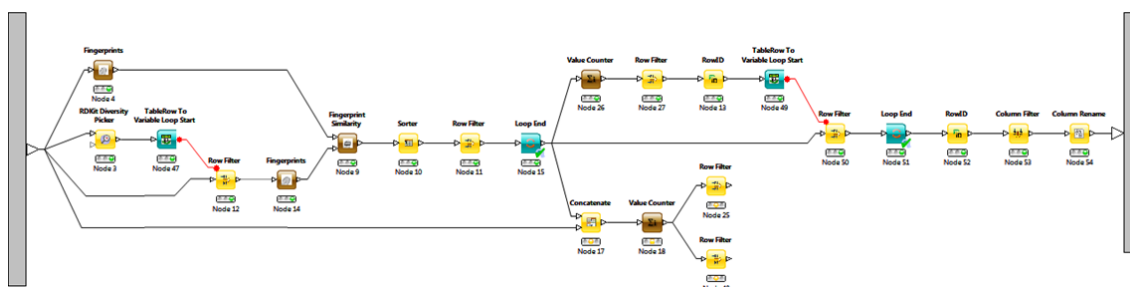


Figure 2.4. Outline of KNIME<sup>242</sup> DiversityCluster Meta Node.

### 2.5.2 Retrieve and React Isothiazolones

Compounds identified by a ChEMBL<sup>159</sup> search against KAT2B were filtered manually and only those with reasonable reported activity against the KAT2B HAT domain ( $< 150 \mu\text{M}$ ) were retained. A KNIME<sup>242</sup> workflow was used to select and sort the isothiazolones from this list and then convert them to the mercaptoacrylamide species, as required for docking.

Of the 110 compounds input, 65 were selected as isothiazolones. These were separated using the Rule-based Row Splitter into those with and without reported  $\text{IC}_{50}$  values. For both subsets the String To Number functionality was used to convert the ChEMBL ID to a number and duplicates then removed using GroupBy ChEMBL ID. Following this, the Affinity was renamed either as  $\text{IC}_{50}$  or as % inhibition, before concatenating to afford one table. This table was then: saved as an SDF file, sorted by  $\text{IC}_{50}$  and saved as a PDF file for printing, and passed to Meta Node ReactToMercaptoacrylamide for conversion of the isothiazolone motif, as required for docking.

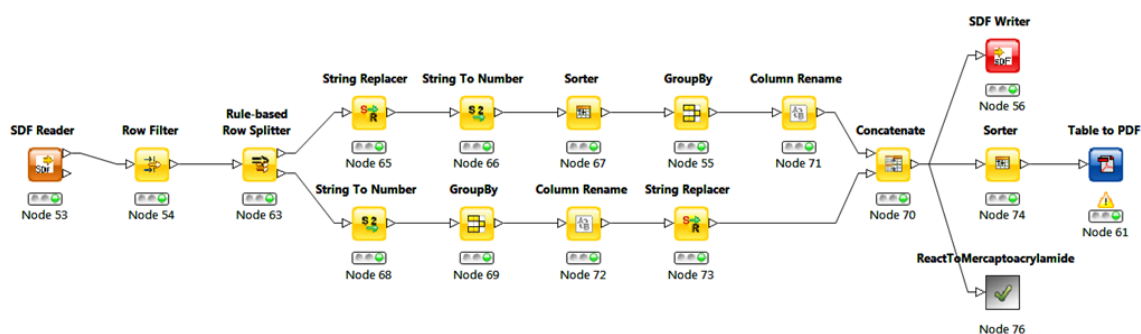


Figure 2.5. Overview of KNIME<sup>242</sup> workflow used to retrieve and react isothiazolones.

#### *ReactToMercaptoacrylamide*

In Meta Node ReactToMercaptoacrylamide the JChem reactor node, JChem 6.1.0, 2013, ChemAxon (<http://www.chemaxon.com>) was used to convert the isothiazolones to the mercaptoacrylamide species, before converting and saving to SDF.

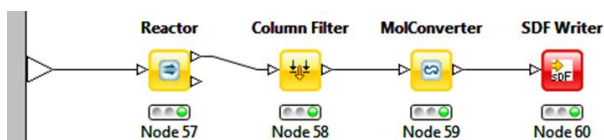


Figure 2.6. Outline of KNIME<sup>242</sup> ReactToMercaptoacrylamide Meta Node.

#### 2.5.3 *Generate Conformers*

A KNIME<sup>242</sup> workflow was designed to generate six conformers of each compound from a library saved in SDF format, give each a unique identifier and output another SDF file.

The list of structures was imported from the SDF file. It was then necessary to convert the structures to a format recognised by the RDKit Add Conformers node. A loop was utilised to generate six conformers for each compound in turn and assign each a unique identifying number. Outside the loop this number was incorporated into the compound reference number, to generate a single unique identifier for each conformer, before saving to SDF.

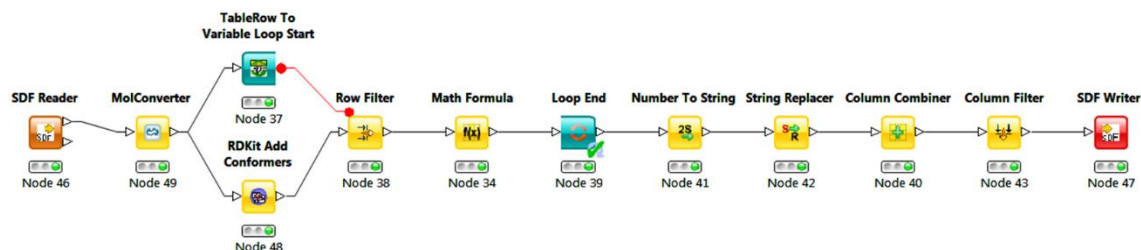
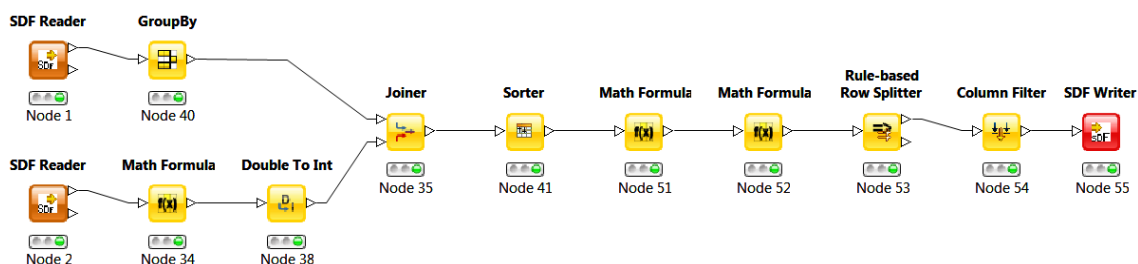


Figure 2.7. Outline of KNIME<sup>242</sup> Workflow Used to Generate Conformers.

#### 2.5.4 *Select Highest Scoring Docking Pose and Retrieve Top 40% of Compounds*

A KNIME<sup>242</sup> workflow was designed to group conformers with multiple poses generated by docking, retaining only the highest scoring pose for each compound, sort these by an associated docking score, retrieve the top scoring 40% and output to an SDF file.

The list of docked structures was imported from an SDF file. The conformers were grouped, retaining the first (best) docking pose and associated score. The resulting data table was joined with that from a compound catalogue, to retrieve the associated catalogue numbers. The data was sorted by ascending docking score and a ranking column was inserted, dependent on the row index. The rows were filtered to retrieve those where  $\text{RANK}/0.4 < \text{ROWCOUNT}$ , i.e. the 40% with the best docking scores. The results were written to an SDF file.

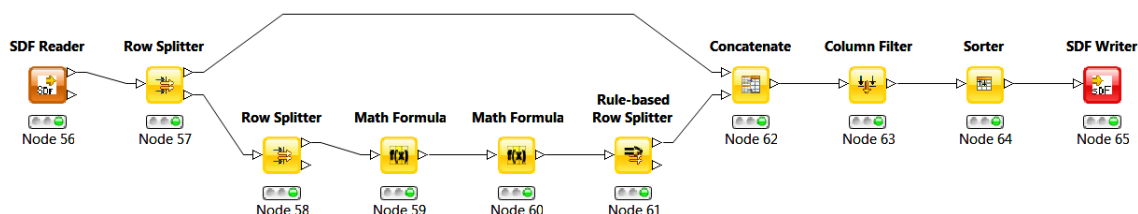


**Figure 2.8. Outline of KNIME<sup>242</sup> Workflow Used to Select Highest Scoring Docked Pose and Retrieve Top 40% of Compounds by Docking Score.**

#### 2.5.5 Generate Final List of Docked Hit Compounds

A KNIME<sup>242</sup> workflow was designed to first split a list of docked compounds according to two columns detailing whether they were interesting because they made new interactions with the protein, or should be ignored because the pose was unreasonable. Those compounds that instigated new interactions were then combined with the top 20% of compounds, by docking score, which did not instigate new interactions but were not unreasonable.

The list of docked structures was imported from an SDF file. The compounds were split depending on whether they formed new interactions. Those that did not form new interactions were split again to remove any marked to be ignored. Those not marked to be ignored were ranked by ascending docking score, dependent on the row index, and the rows were filtered to retrieve those where  $\text{RANK}/0.2 < \text{ROWCOUNT}$ , i.e. the 20% with the best docking scores. The top scoring 20% were combined with the compounds that formed new interactions to generate a final hit list, which was sorted by ascending docking score and saved to SDF.



**Figure 2.9. Outline of KNIME<sup>242</sup> Workflow Used to Split List of Compounds According to Two Columns and Partially Recombine to Generate Final List of Hits.**

## 2.6 Data Analysis and Visualisation

Instant JChem was used for structure database management and searching, JChem 6.1.0, 2013, ChemAxon (<http://www.chemaxon.com>).

Unless otherwise specified, numerical data was analysed using Microsoft Office Excel (2013) and GraphPad Prism version 7.02<sup>243</sup> for Windows, GraphPad Software, La Jolla California USA, [www.graphpad.com](http://www.graphpad.com), in combination.

2D representations of chemical structures and reactions were generated using ChemDraw® Prime version 15.1, PerkinElmer, Waltham, Massachusetts USA, [www.perkinelmer.co.uk](http://www.perkinelmer.co.uk).

## 2.7 Molecular Docking

Ligand:Receptor docking and protein visualisation was performed using MOE.<sup>187</sup> Forcefield Amber10:EHT and Solvation R-Field were employed as standard. The Structure Preparation function was used to prepare, correct and protonate protein structures, which were then Energy Minimised. The Site Finder functionality was used to detect structural pockets available for ligand binding, dummies were created at the alpha sphere centres, and a molecular surface was generated within 4.5 Å of these dummy atoms using the Surfaces and Maps functionality, to enable visualisation of the pocket.

Ligand databases, stored in SDF format, were imported and converted to MDB. The compounds were Energy Minimised using the Calculate Partial Charges, Add Hydrogens and Preserve Existing Chirality options. Pharmacophores were generated using the Pharmacophore Editor.

### 2.7.1 KAT2A HAT Domain

The isothiazolones and corresponding mercaptoacrylamides were docked into the structure of the *Homo sapiens* KAT2A HAT domain (PDB ID: 1Z4R)<sup>241</sup> using the following options:

Protocol: Induced Fit  
Receptor: Receptor+Solvent  
Site: Dummy Atoms  
Pharmacophore: PH4 File  
Ligand: MDB File  
Placement: Alpha Triangle  
Rescoring 1: London dG  
Retain: 30  
Refinement: Forcefield  
Rescoring 2: GBVI/WSA dG  
Retain: 30

### 2.7.2 KAT2B Bromodomain

The amides and sulphonamides were docked into the structure of the *Homo sapiens* PCAF BRD (PDB ID: 5FE8)<sup>240</sup> using the following options:

Protocol: Rigid Receptor  
 Receptor: Receptor+Solvent  
 Site: Dummy Atoms  
 Pharmacophore: PH4 File  
 Template: Selected Atoms  
 Ligand: MDB File  
 Placement: Template Forcing  
 Rescoring 1: London dG  
 Retain: 30  
 Refinement: Forcefield  
 Rescoring 2: GBVI/WSA dG  
 Retain: 30

## 2.8 Protein Techniques

### 2.8.1 Transformation and Expression

#### *SMC5 Head Domain Protein*

N-terminal His6-tagged SMC5 head domain protein was derived from construct AA5, comprised of the N- and C-terminal regions of SMC5 joined by a short linker.

Harvested cells expressing His6-SMC5 head domain protein were generously provided by Dr Aaron Alt (Genome Damage and Stability Centre, University of Sussex).

#### *KAT2A BRD Protein*

His6-tagged KAT2A BRD construct was generously provided by Dr Antony Oliver (Genome Damage and Stability Centre, University of Sussex).

---

Expressed MHHHHHSSGVDLGTENLYFQSMELKDPDQLYTTLKNLLAQIKSHPSAWPFMEPVKKSEAPDY  
 Sequence VIRFPIDLKTMTTERLRSRYVTRKLFVADLQRVIANCREYNPPDSEYCRCSALEKFFYFKLKEGGLIDK

---

The construct was transformed into competent BL21(DE3) cells (New England Biolabs Inc., Massachusetts, USA). 1 µL plasmid was added to 50 µL competent cells and they were incubated for 30 min on ice. The cells were transferred to a heating block for 45 s at 42 °C and returned to ice. 200 µL SOC medium (69319) (Novagen, Merck KGaA, Darmstadt, Germany) was added and the cells were incubated for 1 h at 37 °C, then spread on to Luria-Bertani Agar (10 g/L tryptone,

5 g/L yeast extract, 10 g/L NaCl, 15 g/L agar, pH 7.5) supplemented with ampicillin and incubated overnight at 37 °C.

The transformed cells were grown in 50 mL of Turbo Broth™ (Molecular Dimensions Ltd, Newmarket, UK) supplemented with ampicillin, for ~4 h at 37 °C until OD<sub>600 nm</sub> > 1.0. This culture was used to inoculate 4 x 1 L of Turbo Broth™ supplemented with ampicillin, and the cells were grown at 37 °C with 220 rpm shaking. Protein expression was induced overnight with 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 20 °C at OD<sub>600 nm</sub> > 1.5. Cultures were harvested by centrifugation (8,100 x g for 15 min at 4 °C) on a Beckman Coulter Avanti J-26 XP Centrifuge (California, USA), and then frozen and stored at -20 °C until later use.

### 2.8.2 Protein Purification

#### *SMC5 Head Domain Protein*

Cells from 6 L of culture were thawed and re-suspended in 50 mL lysis buffer (50 mM HEPES pH 7.5, 250 mM NaCl, 10 mM imidazole and 0.5 mM TCEP, supplemented with cOmplete protease inhibitor tablets (Roche Diagnostics Ltd, Burgess Hill, UK) and 10 µL benzonase), filtered using a Millex-GP syringe-driven filter unit, 0.22 µm pore size (Millipore Corporation, MA, USA). Cells were lysed by sonicating for 5 min, using a VCX Vibra-Cell™ ultrasonic processor (Sonics & Materials Inc., Connecticut, USA). The lysate was cleared by centrifugation (37,000 x g for 1 h at 4 °C) on a Beckman Coulter Allegra™ 64R Centrifuge (California, USA). The supernatant was incubated with Talon® His-Tag Purification Resin (TaKaRa Bio, Saint-Germain-en-Laye, France), pre-equilibrated with lysis buffer, for 30 min at 4 °C, before washing with six column volumes of lysis buffer. The His6-tagged protein was eluted from the column using lysis buffer with 300 mM imidazole. Fractions were collected and monitored by SDS-PAGE using TruPAGE™ Precast Gels (Sigma-Aldrich Company Ltd., Dorset, UK) using gel run conditions of 200 V, 40 min, in TruPAGE™ Tris-MOPS SDS Express Running Buffer (Sigma-Aldrich Company Ltd., Dorset, UK). The appropriate fractions were recombined and the protein was concentrated to a volume of 10 mL using Vivaspin® 20 (10 kDa MWCO) centrifugal concentrators (Generon, Slough, UK). The His6-tagged protein was further purified by size exclusion chromatography on a Superdex 200 size exclusion column (GE Healthcare, Illinois, USA), using a storage buffer of 20 mM HEPES pH 7.5, 250 mM NaCl and 0.5 mM TCEP. The fractions were examined by SDS-PAGE and recombined appropriately. Finally, the protein was re-concentrated to 12 mg/mL, flash frozen in liquid nitrogen and stored at -80 °C until later use. Protein concentration was estimated using a NanoDrop™ 2000c Spectrophotometer (Thermo Fisher Scientific, Hertfordshire, UK).

*KAT2A BRD Protein*

Cells from 4 L of culture were thawed and re-suspended in 50 mL lysis buffer (see Table 2.14, supplemented with cOmplete protease inhibitor tablets (Roche Diagnostics Ltd, Burgess Hill, UK)), filtered using a Steritop™ Filter Unit, 0.22 µm pore size (Merck KGaA, Darmstadt, Germany). Cells were lysed by sonicating for 5 min, using a VCX Vibra-Cell™ ultrasonic processor (Sonics & Materials Inc., Connecticut, USA). The lysate was cleared by centrifugation (40,000 x g for 1 h at 4 °C) on a Beckman Coulter Allegra™ 64R Centrifuge (California, USA). The supernatant was filtered using a Minisart® syringe-driven filter unit, 5 µm pore size (Sartorius Stedim Biotech, Aubagne, France), and incubated with Talon® His-Tag Purification Resin (TaKaRa Bio, Saint-Germain-en-Laye, France), pre-equilibrated with lysis buffer, for 30 min at 4 °C, before washing with twenty column volumes of lysis buffer. The His6-tagged protein was eluted from the column using lysis buffer with 300 mM imidazole. Fractions were collected and monitored by SDS-PAGE using TruPAGE™ Precast Gels (Sigma-Aldrich Company Ltd., Dorset, UK) using gel run conditions of 200 V, 40 min, in TruPAGE™ Tris-MOPS SDS Express Running Buffer (Sigma-Aldrich Company Ltd., Dorset, UK). The appropriate fractions were recombined and if necessary the protein was treated overnight at 4 °C with TEV protease to remove the hexa-histidine expression tag, Table 2.14. The protein was concentrated to a volume of 10 mL using Vivaspin® 20 (5 kDa MWCO) centrifugal concentrators (Generon, Slough, UK) and further purified by size exclusion chromatography on a Superdex 75 size exclusion column (GE Healthcare, Illinois, USA), using storage buffer, see Table 2.14. The fractions were examined by SDS-PAGE and recombined appropriately. Finally, the protein was re-concentrated to ~10 mg/mL and stored at -80 °C until later use. Protein concentration was estimated using a NanoDrop™ 2000c Spectrophotometer (Thermo Fisher Scientific, Hertfordshire, UK).

**Table 2.14. Purification Details for KAT2A BRD Protein.**

Purpose	Lysis Buffer	Storage Buffer	His6-Tag Cleaved
Tm Shift	50 mM HEPES pH 7.5, 250 mM NaCl, 10 mM imidazole and 0.5 mM TCEP	20 mM HEPES pH 7.5, 250 mM NaCl and 0.5 mM TCEP	N
NMR	50 mM Sodium Phosphate pH 7.5, 500 mM NaCl, 10 mM imidazole and 0.5 mM TCEP	20 mM Sodium Phosphate pH 7.5, 500 mM NaCl, and 0.5 mM TCEP	Y
ITC	50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM imidazole and 0.5 mM TCEP	20 mM HEPES pH 7.5, 500 mM NaCl and 0.5 mM TCEP	Y

**2.9 Screening Compounds**

Unless otherwise specified, small molecule compounds and fragments were prepared by colleagues in the Sussex Drug Discovery Centre (University of Sussex), or obtained from in-house

libraries, as solid samples or dissolved in DMSO at 10 mM (compounds) or 100 mM (fragments). The in-house fragment library is comprised of the Maybridge Ro3 2500 Diversity Fragment Library (Maybridge, Altrincham, UK) and the Life Chemicals 3D Fragments Library (Life Chemicals Inc., Niagara-on-the-Lake, Canada). Additionally, a library of quinolinone amides and sulphonamides was prepared by David Pearce (Automated High-Throughput Medicinal Chemistry and Synthesis Lab, University of Sussex). Other chemicals were of analytical grade and obtained from Sigma-Aldrich Company Ltd. (Dorset, UK).

## 2.10 Isothermal Titration Calorimetry

### 2.10.1 SMC5 Head Domain Protein

Purified SMC5 head domain protein, stored at -80 °C, was thawed rapidly and dialysed into a buffer of 20 mM HEPES pH 7.5, 150 mM NaCl and 5 mM MgCl, using dialyzer tubes, D-Tube™ Dialyzer Midi, MWCO 3.5 kDa (Novagen, Merck KGaA, Darmstadt, Germany) and the protein concentration measured using a NanoDrop™ 2000c Spectrophotometer (Thermo Fisher Scientific, Hertfordshire, UK).

ITC experiments were carried out on a MicroCal ITC200 (GE Healthcare, Illinois, USA), with a cell volume of 200 µL, and were performed at 30 °C in 20 mM HEPES pH 7.5, 150 mM NaCl and 5 mM MgCl, with the pH of the compound solutions adjusted by addition of Tris HCl pH 9.5 or HEPES pH 8.0. All titrations were conducted using an initial injection of 0.9 µL followed by ten identical injections of 3.8 µL at 240 s intervals. The dilution heats were measured in separate experiments and were subtracted from the titration data. The data was analysed using the MicroCal PEAQ-ITC Analysis Software. The compound and protein concentrations employed are quoted alongside the resulting data.

### 2.10.2 KAT2A BRD Protein

Purified KAT2A bromodomain protein, stored at -80 °C, was thawed rapidly and dialysed into a buffer of 20 mM HEPES pH 7.5 and 150 mM NaCl, using BioDesignDialysis Tubing™, MWCO 3.5 kDa (BioDesign Inc. of New York, USA) or Slide-A-Lyzer™ Dialysis Cassettes, MWCO 3.5 kDa (Thermo Fisher Scientific, Hertfordshire, UK), and the protein concentration measured using a NanoDrop™ Lite Spectrophotometer (Thermo Fisher Scientific, Hertfordshire, UK).

ITC experiments were carried out on a MicroCal ITC200 (GE Healthcare, Illinois, USA) or a MicroCal PEAQ-ITC (Malvern Instruments, Worcestershire, UK), with cell volumes of 200 µL, and were performed at 30 °C (ITC200) or 25 °C (PEAQ-ITC) in 20 mM HEPES pH 7.5 and 150 mM NaCl. Titrations were conducted using an initial injection of 0.9 µL followed by ten identical injections of 3.8 µL at 240 s intervals (ITC200) or 0.4 µL followed by twelve identical injections of 3.0 µL at



150 s intervals (PEAQ-ITC). The dilution heats were measured in separate experiments and were subtracted from the titration data. The data was analysed using the MicroCal PEAQ-ITC Analysis Software and the single site binding model was employed to determine binding constants. The compound and protein concentrations employed are quoted alongside the resulting data.

## 2.11 HAT Activity Assays

### 2.11.1 Fluorescence Based Activity Assay

Experiments with KAT2A were performed using the N-terminal GST-tagged form (KAT2A (GCN5), Active (K311-381G)) and histone H3 Peptide (1-21) (H12-58-500) from SignalChem (BC, Canada). Acetyl coenzyme A sodium salt, Fluorescein diacetate 5-maleimide and ThioGlo-5 were purchased from Sigma-Aldrich Company Ltd. (Dorset, UK), and Thiol Fluorescent Probe IV from Merck Millipore (Billerica, MA, USA). All other chemicals were of analytical grade and obtained from Sigma-Aldrich Company Ltd. (Dorset, UK).

All experiments were conducted in black FLUOTRAC™ 384 well plates (781076) (Greiner Bio-One, Kremsmünster, Austria) in a buffer of 50 mM HEPES, 0.1 mM EDTA, 0.01% Tween-20, pH 7.5, with a final reaction volume of 40  $\mu$ L. Individual assay components are listed alongside the resulting experimental data. Unless otherwise specified, when measuring enzyme activity, the reaction was initiated by addition of a mixture of the acetyl coenzyme A, histone H3 peptide and fluorescence indicator to the KAT2A. Fluorescence was monitored using a FlexStation® 3 microplate reader (Molecular Devices, California, USA) with excitation and emission wavelengths as specified in Table 2.15.

**Table 2.15. Excitation and Emission Wavelengths for Fluorescence Indicators.**

Fluorescence indicator	$\lambda_{\text{ex}}$ (nm)	$\lambda_{\text{em}}$ (nm)	Cutoff (nm)
Fluorescein diacetate 5-maleimide	340	540	530
Thiol Fluorescent Probe IV	400	465	455
ThioGlo-5	285	515	515

### 2.11.2 EpiQuik™ Activity Assay

All experiments were performed using GCN5L2 generously provided by the Structural Genomics Consortium (Toronto, Canada). The EpiQuik™ HAT activity assay kit (P-4003) was purchased from Epigentek (Farmingdale, NY, USA). All other chemicals were of analytical grade and obtained from Sigma-Aldrich Company Ltd. (Dorset, UK).

HAT activity was measured using the EpiQuik™ calorimetric assay kit and the absorbance was read on a PHERAstar® FS plate reader (BMG Labtech, Ortenberg, Germany) at 450 nm.

## 2.12 KAT2A BRD T<sub>m</sub> Shift Assay

Experiments were performed with KAT2A BRD with and without a His6-tag, prepared in-house. SYPRO® Orange Protein Gel Stain, 5000x concentrate (S5692), was purchased from Sigma-Aldrich Company Ltd. (Dorset, UK) and DMSO (D/4120/PB08) from Fisher Chemical (Thermo Fisher Scientific, Hertfordshire, UK). All other chemicals were of analytical grade and obtained from Sigma-Aldrich Company Ltd. (Dorset, UK).

All experiments were conducted in white LightCycler® 480 96 well plates (04 729 692 001) (Roche Molecular Systems, CA, USA), with a final reaction volume of 20 µL. Buffers are specified in Table 2.16. Individual assay components are listed alongside the resulting experimental data. Post-optimisation, experiments were conducted at 5x SYPRO® Orange and 2 µM KAT2A BRD. SYPRO® Orange was added to a solution of the protein with or without 1 mM test compound and 2% DMSO and the plate was sealed using a clear LightCycler® 480 sealing foil (04729757001) (Roche Molecular Systems, CA, USA). The plate was equilibrated briefly on ice, before measuring the fluorescence whilst raising the temperature from 20 °C to 85 °C, using the LightCycler® 480 Instrument, 96 well (05 015 278 001) (Roche Molecular Systems, CA, USA) with excitation and emission wavelengths 465 nm and 580 nm respectively. The protein melting temperature (T<sub>m</sub>) was determined using the Roche Protein Melting Analysis Software and T<sub>m</sub> shift values calculated by comparing the T<sub>m</sub> of the protein with and without addition of compound.

**Table 2.16. Buffer Details for KAT2A BRD T<sub>m</sub> Shift Assays.**

KAT2A BRD Species	Buffer
His6-KAT2A BRD	10 mM HEPES pH 7.5 and 100-700 mM NaCl
KAT2A BRD	20 mM Sodium Phosphate pH 7.5 and 200 mM NaCl

## 2.13 X-Ray Crystallography

### 2.13.1 Crystal Growing in Hanging Drops

His6-KAT2A bromodomain protein was predominantly crystallised in hanging drops, with 0.1 M Tris, pH 8.5 and 16% PEG 3350 at 5 mg/mL protein at 4 °C. 24 well VDX™ plates, pre-applied with sealant (HR3-170) (Hampton Research, California, USA), were prepared with 1 mL reservoir solution of 0.1 M Tris, pH 7.5 or pH 8.5 and 16%-20% PEG 3350. 1 µL protein, stored at -80 °C, was transferred to siliconized cover slips, 22 mm round (MD4-04) (Molecular Dimensions Ltd, Newmarket, UK) and 1 µL reservoir solution was added. The cover slips were inverted and pressed on to the wells to form a seal. The plates were incubated at 4 °C and crystal growth was monitored using a Leica M165 C microscope (Leica Microsystems, Wetzlar, Germany).

### 2.13.2 Crystal Soaking

Ligand-bound structures were obtained predominantly via crystal soaking. Suitable crystals, selected dependent on their size and morphology, were soaked for 1 h at 4 °C in a 10 µL drop of reservoir solution supplemented with test compound, Table 2.17.

**Table 2.17. Crystal Soaking Conditions.** Soaking conditions used to obtain ligand-bound structures of KAT2A BRD.

Compound	Compound Soaking Conditions
UOS-21430	10 mM with 10% DMSO
UOS-18715	100 mM with 20% Methanol
UOS-28438	10 mM with 10% DMSO
UOS-28625	10 mM with 10% DMSO
UOS-28734	10 mM with 10% DMSO
UOS-27986	100 mM with 20% DMSO
UOS-31388	10 mM with 10% DMSO

### 2.13.3 Co-Crystallisation

UOS-31181-bound His6-KAT2A BRD was crystallised in sitting drops, with 0.1 M sodium acetate, pH 5.0 and 1.5 M ammonium sulphate at 6 mg/mL protein at 20 °C. UOS-31181 was added to His6-KAT2A BRD to afford a protein-ligand solution with 10 mM ligand at 33% DMSO (D/4120/PB08) (Fisher Chemical, Thermo Fisher Scientific, Hertfordshire, UK). MRC 2 drop well 96 well crystallization plates, manufactured by Swissci (HR3-107) (Hampton Research, California, USA), were prepared using a Crystal Phoenix protein crystallography dispenser (Art Robbins Instruments, California, USA). Four 96-condition crystallisation screens were conducted with the JCSG-*plus*<sup>™</sup> HT-96 (MD1-40), ProPlex HT-96 (MD1-42), Morpheus<sup>™</sup> (MD1-47) and PACT *premier*<sup>™</sup> HT-96 (MD1-36) commercial libraries (Molecular Dimensions Ltd, Newmarket, UK). 50 µL reservoir solution was dispensed into each well and 0.15 µL protein-ligand solution into each drop well, to which 0.15 µL reservoir solution was added. Plates were sealed with ClearVue Sealing Sheets (MD6-01S) (Molecular Dimensions Ltd, Newmarket, UK) and incubated at 20 °C. Crystal growth was monitored using a Leica M165 C microscope (Leica Microsystems, Wetzlar, Germany).

### 2.13.4 Diffraction Data Collection, Structure Solution and Refinement

Crystals of suitable size and morphology were cryoprotected by successive soaking in reservoir solution supplemented with 10%, 20% and 30% ethylene glycol, mounted on to CrystalCap<sup>™</sup> SPINE HT CryoLoops (Hampton Research, California, USA), and flash frozen in liquid nitrogen or examined immediately.

**Table 2.18. Data Collection and Refinement Statistics.** Statistics from single crystals of apo or ligand-bound KAT2A BRD. Averaged data presented alongside that for highest shell (in parentheses).

	KAT2A Apo	21430	18715	28438	28625	28734	27986	31388	31181
<b>Data Collection</b>									
Space Group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	F432
a (Å)	45.89	45.73	45.838	45.804	45.644	45.756	45.783	45.816	161.76
b (Å)	73.38	73.43	72.96	73.428	73	72.995	73.211	73.336	161.76
c (Å)	77.24	77.33	76.746	77.396	77.189	77	77.367	77.208	161.76
α (°)	90	90	90	90	90	90	90	90	90
β (°)	90	90	90	90	90	90	90	90	90
γ (°)	90	90	90	90	90	90	90	90	90
Wavelength (Å)	0.97625	0.97625	1.5418	1.5418	1.5418	1.5418	1.5418	1.5418	1.5418
Resolution Limit (Å)	53.2-1.53 (1.57-1.53)	73.4-1.54 (1.58-1.54)	39.4-2.0 (2.07-2.00)	53.3-2.0 (2.07-2.00)	53.0-2.0 (2.07-2.00)	53.0-1.7 (1.76-1.70)	53.2-2.2 (2.28-2.20)	39.4-2.2 (2.28-2.20)	57.2-1.8 (1.86-1.80)
Unique Observations	40019 (2891)	39094 (2868)	17962 (1695)	18125 (1690)	17937 (1678)	28996 (2821)	13726 (1339)	13681 (1308)	17357 (1663)
Completeness (%)	99.8 (99.9)	99.5 (99.5)	99.6 (96.4)	99.3 (95.3)	99.2 (94.8)	99.7 (98.6)	99.9 (99.6)	99.5 (98.2)	99.9 (99.2)
Multiplicity	4.7 (4.7)	4.5 (3.9)	4.8 (4.6)	4.9 (4.8)	4.8 (3.0)	9.1 (3.6)	4.9 (4.9)	4.5 (4.4)	8.2 (4.6)
Rmerge %	0.116 (0.945)	0.083 (0.818)	0.142 (0.494)	0.186 (0.616)	0.134 (0.418)	0.100 (0.463)	0.138 (0.381)	0.264 (0.674)	0.065 (0.692)
Rpim(l) %	0.065 (0.540)	0.050 (0.544)							
CC <sub>1/2</sub>	0.993 (0.545)	0.996 (0.539)							
I/σ	6.8 (1.4)	8.0 (1.2)	5.5 (1.9)	4.5 (1.8)	5.9 (1.9)	12.2 (2.0)	6.4 (2.7)	3.5 (1.1)	15.1 (1.0)
<b>Refinement</b>									
Resolution Range (Å)	22.4-1.53	22.3-1.54	76.7-2.00	53.3-2.0	53.0-2.0	53.0-1.70	53.2-2.2	53.2-2.20	93.4-1.80
R <sub>cryst</sub>	0.179	0.188	0.219	0.221	0.211	0.181	0.223	0.264	0.235
R <sub>free</sub>	0.2	0.21	0.262	0.276	0.269	0.229	0.28	0.321	0.268
Mean B	24.2	28.1	27.29	31.2	28.73	25.24	27.8	36.33	32.99
RMSD bond length (Å)	0.01	0.01	0.017	0.017	0.019	0.021	0.018	0.013	0.014
RMSD bond angle (°)	0.84	0.86	1.739	1.751	1.939	1.938	1.831	1.575	1.789

X-ray diffraction data was generously collected and analysed by Dr Mark Roe (X-Ray Crystallography Collaborative Research Facility, University of Sussex). The data was collected on the in-house X-ray system, consisting of a Rigaku MicroMax™-007 HF generator, VariMax-HF mirrors, Saturn 944+ CCD detector and Oxford Cryosystems 700 cryostream, or using synchrotron radiation at the Diamond Light Source on beamline I03. The diffraction images were indexed and integrated, the data scaled and averaged and the structures solved and refined. Data collection and refinement statistics were provided by Dr Mark Roe, Table 2.18.

PyMOL 1.6<sup>232</sup> was used to produce 3D representations of the protein structures.

## 2.14 KAT2A BRD TR-FRET Assay

Experiments were performed using N-terminal GST-fusion KAT2A bromodomain (RD-11-258), from CisBio (Bedford, MA, USA). Ligand-biotin compounds UOS-31490 and UOS-31670 were generously synthesised by Dr Lewis Pennicott (Sussex Drug Discovery Centre, University of Sussex). DMSO (D/4120/PB08) was purchased from Fisher Chemical (Thermo Fisher Scientific, Hertfordshire, UK). All other chemicals were of analytical grade and obtained from Sigma-Aldrich Company Ltd. (Dorset, UK).

All experiments were conducted in white LUMITRAC™ 384 well low volume plates (784075) (Greiner Bio-One, Kremsmünster, Austria), at 1% DMSO concentration, with a final reaction volume of 20 µL and were incubated at room temperature. Fluorescence was monitored using a PHERAstar® FS plate reader (BMG Labtech, Ortenberg, Germany) with an HTRF® appropriate optic module, employing a 60 µs integration delay, 400 µs integration time and 300 flashes, with excitation wavelength 370 nm and emission wavelengths 620 nm and 665 nm. Results were expressed as HTRF® ratios, and negative controls, without GST-KAT2A, were used to assess non-specific signal, to calculate the assay window and specific signal.

- $\text{HTRF}^{\circ} \text{ ratio} = (665 \text{ nm}/620 \text{ nm}) \times 10^4$
- $\text{Assay window} = \text{HTRF}^{\circ} \text{ ratio positive} / \text{HTRF}^{\circ} \text{ ratio negative}$
- $\text{Specific signal, HTRF}^{\circ} \Delta \text{ ratio} = \text{HTRF}^{\circ} \text{ positive} - \text{HTRF}^{\circ} \text{ negative}$

The data was analysed using either Microsoft Office Excel (2013) and GraphPad Prism version 7.02<sup>243</sup> for Windows, GraphPad Software, La Jolla California USA, [www.graphpad.com](http://www.graphpad.com), or the Studies and Vortex packages from Dotmatics Suite version 5.1<sup>244</sup>, Dotmatics Limited, Bishops Stortford UK, [www.dotmatics.com](http://www.dotmatics.com).

### 2.14.1 Assay Design

The EPIgeneous™ Binding Domain Discovery Kit (62BDDPEG) (Cisbio, Bedford, MA, USA) was utilised in assay design, to identify a suitable biotinylated ligand and determine the optimal combination of donor and acceptor dyes. Titrations of UOS-31490 or UOS-31670 and streptavidin -acceptor (SA-d2 or SA-XL665) at a ligand-biotin : SA-acceptor ratio of 8:1 were conducted as per the kit instructions, with 5 nM GST-KAT2A BRD and 1x anti-GST-donor cryptate ( $\text{Eu}^{3+}$  or  $\text{Tb}^{2+}$ ), using Binding Domain Diluent Buffer, Binding Domain Detection Buffer #1 and Binding Domain Detection Buffer #2 as appropriate. The plates were incubated for three hours. Three detector reagent pairings were assessed, anti-GST- $\text{Eu}^{3+}$  with SA-d2, anti-GST- $\text{Eu}^{3+}$  with SA-XL665 and anti-GST- $\text{Tb}^{2+}$  with SA-XL665.

### 2.14.2 Assay Optimisation and Screening

In assay optimisation and screening the interaction between GST-KAT2A BRD and ligand-biotin compound UOS-31670 was detected using anti-GST- $\text{Eu}^{3+}$  (61GSTKLA) and SA-XL665 (610SAXLA) using Binding Domain Diluent Buffer (62DLBDDF) and Binding Domain Detection Buffer #1 (62DB1FDG), all purchased from Cisbio (Bedford, MA, USA).

For optimisation, assay conditions are listed alongside the resulting experimental data. In screening, 2.5 nM GST-KAT2A BRD, 80 nM UOS-31670, 0.313 ng anti-GST- $\text{Eu}^{3+}$  and 10 nM SA-XL665 concentrations were used. GST-KAT2A BRD in 4  $\mu\text{L}$  Binding Domain Diluent Buffer was added using a Multidrop™ Combi Reagent Dispenser (Thermo Fisher Scientific, Hertfordshire, UK), to 2  $\mu\text{L}$  compound in Binding Domain Diluent Buffer with 10% DMSO previously dispensed using an XPP-721 Automated Liquid Handler (FluidX, Manchester, UK). After 15 min, 14  $\mu\text{L}$  of a solution of 4:5:5 UOS-31670 in Binding Domain Diluent Buffer : anti-GST- $\text{Eu}^{3+}$  in Binding Domain Detection Buffer #1 : SA-XL665 in Binding Domain Detection Buffer #1 was added using the Multidrop™ Combi Reagent Dispenser. The plates were incubated for two hours before reading. Each compound was assayed in at least two independent experiments.

For single point screens, compounds were tested at 1 mM final assay concentration and percentage inhibition was calculated relative to positive and negative controls on a per plate basis. For  $\text{IC}_{50}$  screens, ten-point dose-response compound titrations were prepared in duplicate. Percentage inhibition was calculated relative to the controls and  $\text{IC}_{50}$  values were determined by fitting the percentage inhibition versus compound concentration, constraining the upper and lower limits to 0% and 100%.

### 3 Assessing the SMC Complexes as Potential Targets for Tool Inhibitor Development

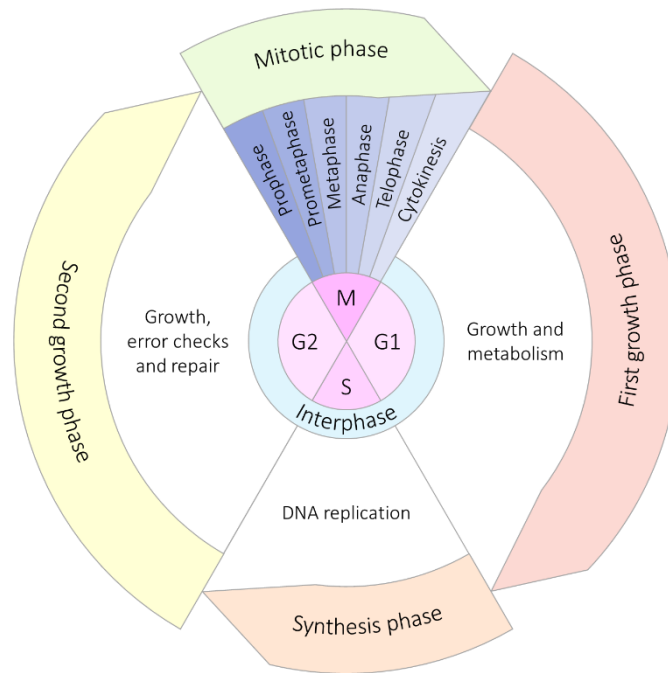
The SMC complexes are critically important in regulating chromosome architecture and organisation and thereby coordinating chromosome condensation, sister chromatid cohesion, DNA repair, homologous recombination and transcriptional regulation. Their biological functions and mechanisms are not yet fully understood, but already they represent interesting oncology targets, particularly due to their roles in DNA repair. As such, they are compelling targets for development of tool inhibitors, which will enable biological investigation and potentially initiation of drug discovery projects.

In this chapter, the druggabilities of the core SMC proteins and NSE subunits of SMC5/6, and cohesin SMC1 and SMC3 were assessed, with the aim to advance to design and characterisation of tool inhibitors for any of the proteins deemed druggable. However, at the time at which this research was completed, the SMC5/6 and cohesin complexes were established as not druggable, within the constraints of accessible resources.

#### 3.1 Introduction

Every human cell contains approximately four meters of DNA that needs to be meticulously packed into the cell nucleus, which has an average diameter of just 6  $\mu\text{m}$ . Moreover, at cell division this DNA must be further compacted into mitotic chromosomes. This compaction is carefully regulated during the cell cycle, Figure 3.1, and is dependent on an array of facilitator proteins, which associate with the DNA to build chromosomes. The structural maintenance of chromosomes (SMC) proteins are principal chromosomal constituents that regulate chromosome architecture and organisation. By virtue of this, they function in a variety of cellular processes including chromosome condensation, sister chromatid cohesion, DNA repair, homologous recombination and transcriptional regulation.<sup>245</sup>

The first SMC protein reported was *Escherichia coli* mukB in 1991.<sup>246</sup> Niki *et al.* observed that mukB mutants exhibit anucleate cells, and used computational analysis to predict that the protein structure comprised of three globular domains interspaced by two  $\alpha$ -helical coiled-coil regions. They also noted that the amino-terminal domain contained a nucleotide binding sequence. Soon after, SMC1 was identified in *Saccharomyces cerevisiae*<sup>247</sup>, and SMC2 and SMC4 in *Xenopus laevis*<sup>248</sup>, where it was also recognised that these associate to form a heterodimer. SMC6, originally known as RAD18, was first identified in *Schizosaccharomyces pombe*.<sup>249</sup> Genes encoding similar SMC proteins have subsequently been found in a wide variety of organisms.<sup>245</sup>



**Figure 3.1. Schematic of the Cell Cycle.**

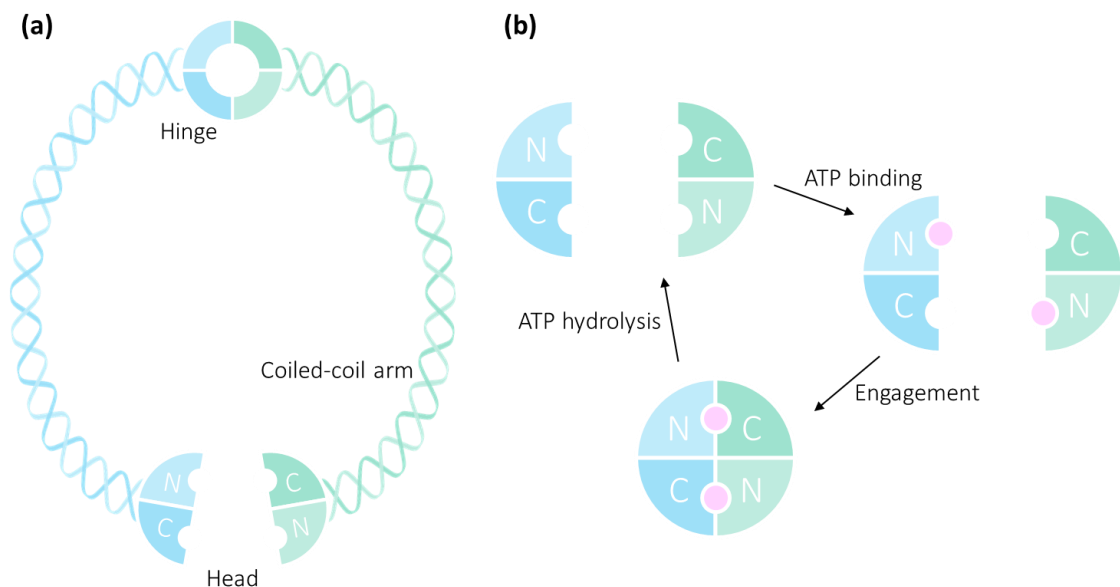
It is now well established that these widely conserved proteins form dimeric structures, which in turn associate with regulatory subunits to form large protein complexes. While it seems most prokaryotes have just one SMC protein that forms a homodimer, eukaryotes have at least six, which interact to form three heterodimers.<sup>250,251</sup> SMC1 and SMC3 are the central components of the cohesin complex, which mediates sister chromatid cohesion and is required for proper chromosome segregation at cell division.<sup>252</sup> SMC2 and SMC4 constitute the core of the condensin complex that is essential for chromosome condensation during mitosis.<sup>253</sup> The third heterodimer is composed of SMC5 and SMC6, whose sequences are substantially divergent from those of SMC1-4. The SMC5/6 complex has multiple roles in DNA repair<sup>254–256</sup> and checkpoint signalling.<sup>257</sup> Interestingly, while the SMC genes are each essential for eukaryotic life, the three heterodimers exhibit partly overlapping roles. For example, cohesin can aid in chromosome condensation<sup>258</sup> and both cohesin and condensin have roles in DNA repair.<sup>259,260</sup>

All SMC proteins share a common architecture. They are composed of about 1,000 amino acids<sup>251</sup>, self-folded by antiparallel coiled-coil interactions to create rod-shaped molecules. The long stretches of coiled-coil are flexible, due to discontinuities<sup>245</sup>, and are connected at one end to an ATP-binding cassette (ABC)-like head domain and at the other to a hinge domain<sup>250</sup>, Figure 3.2 (a). Nucleotide-binding Walker A and Walker B motifs reside separately in the N- and C-termini respectively, which together compose the globular ATPase head domain.<sup>251</sup> Two SMC subunits interact tightly at their hinge domains to produce a V-shaped heterodimer.<sup>261</sup> ATP binding at the SMC heads then promotes formation of the characteristic ring-shaped dimer. The head domains form nucleotide-sandwiched dimers in which two molecules of ATP reside at the



interface, each bound via  $\alpha$  and  $\beta$  phosphates to the Walker A motif of one SMC subunit and via the  $\gamma$  phosphate to the Walker B motif of the second.<sup>224</sup> Engagement and disengagement of the ATPase heads is controlled by binding and hydrolysis of ATP, which is thought to play a vital role in the dynamic interactions of SMC proteins and DNA<sup>262</sup>, Figure 3.2 (b).

In addition to the ATP-dependent binding, the SMC heads are held together by 'kleisin' subunits.<sup>245</sup> These kleisins are less well conserved than the SMCs. They make asymmetric contacts with the SMC proteins, with the N-terminus forming a helix bundle with the coiled-coil adjacent to one SMC head domain, while a winged-helix domain at the C-terminus interacts with the bottom surface of the ATPase domain of the other.<sup>263,264</sup> Additional non-conserved subunits assemble around the kleisin.<sup>245</sup> The bridging kleisin can cause the SMC complexes to appear as closed ring structures by electron microscopy.<sup>265</sup> As might be implied by this shape, according to the 'embrace' model, these complexes bind DNA by topological embrace. There is evidence that the eukaryotic SMC complexes are linked to minichromosomes *in vivo* and it seems that topological entrapment is essential for complex function.<sup>266–268</sup> This does not preclude additional functions not reliant on topological linking.<sup>245</sup>



**Figure 3.2. Structure of SMC Heterodimers.** Each SMC subunit is self-folded by antiparallel coiled-coil interactions to create a rod-shaped molecule with a hinge domain at one end and an ATP-binding head domain at the other. (a) Two SMC subunits interact tightly at their hinge domains to produce a V-shaped heterodimer. (b) ATP binding at the SMC heads drives formation of nucleotide-sandwiched dimers with two molecules of ATP at the interface. Engagement and disengagement of the ATPase heads is controlled by binding and hydrolysis of ATP. Figure adapted from Losada & Hirano, 2005.<sup>250</sup>

Most of the time SMC complexes are in dynamic contact with chromosomes, with the notable exception of cohesin on sister chromatids following DNA replication.<sup>269</sup> If the complexes function by topological binding to DNA, this provokes the question of how the DNA enters and exits the ring. In attempt to identify the mechanism of cohesin loading and unloading on DNA, individual

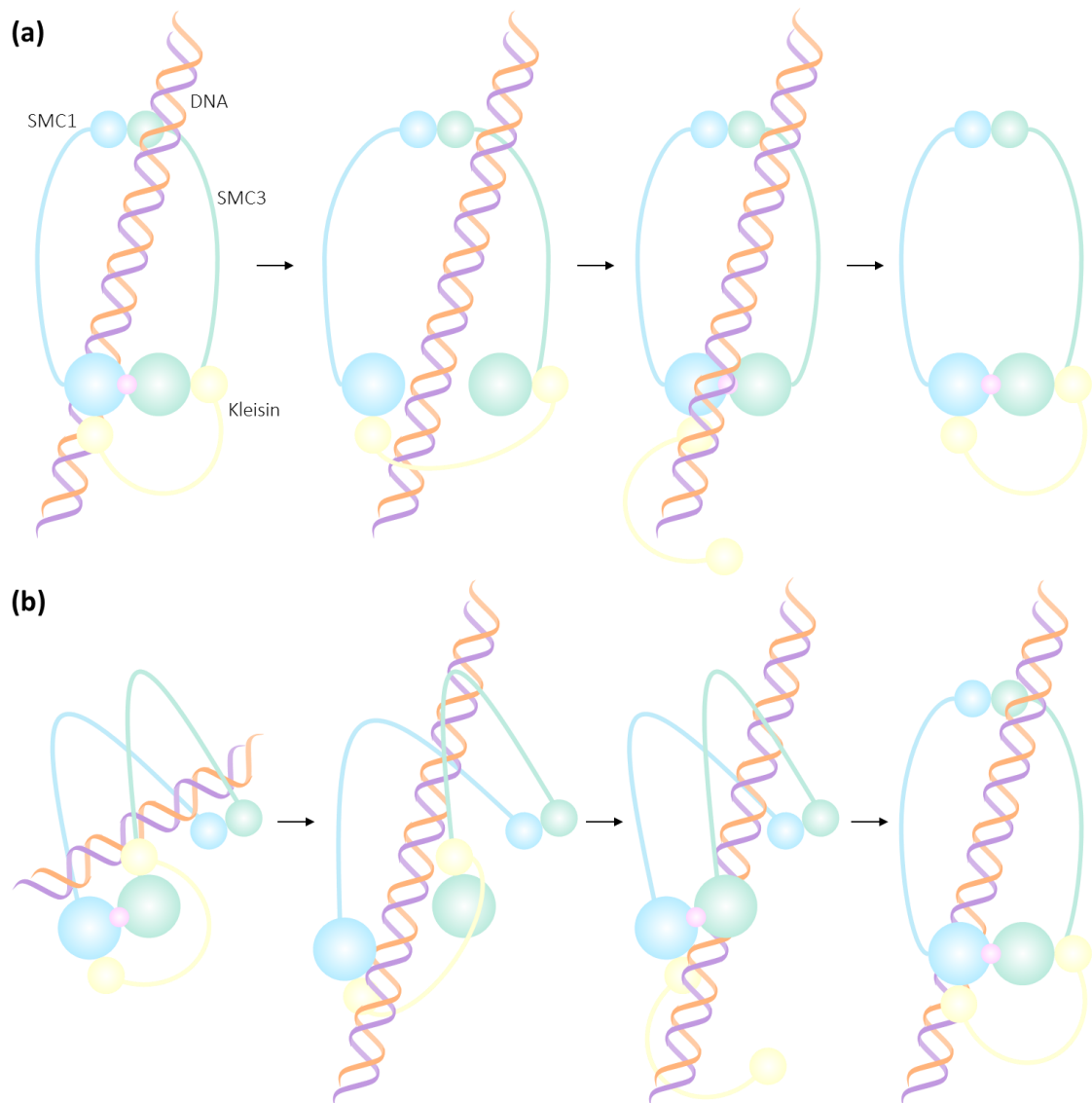
subunit interfaces can be sealed by covalent linkage or ligand-induced dimerisation.<sup>245</sup> Using this approach it was ascertained that DNA loading is dependent only on the SMC hinge domain interface, which would suggest that DNA enters via the cohesin hinge.<sup>270,271</sup> Contrastingly, locking the SMC3-kleisin interface stabilised the cohesin on the chromosome, suggesting a role for this interface in DNA exit.<sup>271</sup> However, crosslinks and ligand-induced dimerization can affect protein function in ways other than the intended interface closures, so these results should be interpreted with caution.<sup>245</sup> A different model has since been proposed, founded on *in vitro* experiments reconstituting cohesin loading on to DNA in *S. pombe*.<sup>272,273</sup> From these it was suggested that loading and unloading follow similar routes, regulated by the tripartite ring structure of the SMC complex, which establishes two sequential interlocking gates. Initially, DNA-sensing lysine residues on the SMC3 head trigger ATP hydrolysis, which promotes the disengagement of the ATPase head, allowing DNA to move through this first barrier. Once the ATP is replenished and therefore the head interaction is re-engaged, the kleisin N-terminus dissociates from SMC3, allowing the DNA to pass fully into or out of the ring. Importantly this model of interlocking gates also explains earlier observations. Covalently linking SMC3 and the kleisin is not sufficient to prevent DNA from interacting with the complex because the DNA is still able to pass through the ATPase gate and thereby interact with cohesin, however, this will impede the exit of any DNA secured within the ring by the kleisin.<sup>273</sup> Contrastingly, if the kleisin is mutated to inhibit binding with SMC3, the cohesin complex is capable of interacting with chromatin but the interaction is unstable, as one of the two gates has been destroyed.<sup>274</sup> The DNA-sensing lysine residues on SMC3 point into the ring, which is ideal for DNA exit but awkward for DNA entry. As DNA is relatively stiff, presumably the flexible cohesin must fold to expose the lysines to initiate loading, Figure 3.3.<sup>245,273</sup> It is likely that the entry and exit of DNA from the other SMC complexes involves similar reactions. Both cohesin and condensin have been observed by atomic force microscopy in the folded conformation suggested for DNA entry<sup>275,276</sup>, and in SMC5/6 non-SMC element 5 (NSE5) and NSE6 have been reported to bind to both the head and hinge, which is possible only if the two come into close proximity.<sup>277,278</sup> This model for DNA loading and unloading however remains just that. There are still many unanswered questions. For example, recent publications suggest that while head engagement is critical for cohesin unloading, ATP hydrolysis is possibly not.<sup>279</sup> Perhaps the SMC-kleisin interaction is less favourable when the SMC heads are engaged and dissociates simply due to steric effects.<sup>280</sup>

SMC complexes appear to concentrate at centromeres and at discrete locations on chromosome arms.<sup>281–283</sup> In *S. cerevisiae* it appears that the accumulation of Cdc7-Dbf4 kinase, facilitated by the COMA (Ctf19-Okp1-Mcm21-Ame1) complex, serves to enrich centromeric cohesin, by

recruitment of the cohesin loader complex.<sup>284</sup> Condensin enrichment is promoted by centromere protector shugoshin.<sup>285</sup> Along the chromosome arms, SMC complex loading occurs most frequently at promoters of highly expressed genes.<sup>286,287</sup> It is not known what defines these sites, although they are often associated with transcription factors.<sup>245</sup> However, it has been shown that the 'remodelling the structure of chromatin' (RSC) complex, which is recruited to open reading frames of actively transcribed genes, in turn recruits the cohesin loader complex.<sup>288</sup> Condensin and cohesin are typically loaded at the same DNA sites.<sup>287</sup> Although condensin is not known to interact with the cohesin loader it appears that this loader complex helps to facilitate its binding<sup>282</sup>, potentially by maintaining nucleosome-free regions at promoter sites.<sup>288</sup> Condensin is also known to interact with TFIIC (transcription factor for polymerase III C) and TATA-box binding protein (TBP), which are involved in recruiting RNA polymerase III to transcribe small untranslated RNAs<sup>289</sup>, and with components of the monopolin complex, which is responsible for maintaining the integrity of ribosomal DNA (rDNA).<sup>290</sup> These interactions can target condensin to tRNA (transfer RNA) genes and rDNA.

After the SMC complexes are loaded it appears that they translocate, retaining topological contact, to reside more permanently at sites of convergent transcriptional termination, between genes that are transcribed in converging directions.<sup>291</sup> Accordingly, in *S. pombe*, the majority of cohesin resides at convergent transcriptional terminators, with a smaller amount coinciding with the cohesin loader.<sup>292</sup> In human cells, cohesin commonly accumulates at sites of the CCCTC-binding factor (CTCF)<sup>293</sup>, the vast majority of which are inter- or intra-genic.<sup>294</sup> In humans, it is unlikely that cohesin will be able to traverse an entire gene before disengaging from the DNA.<sup>245</sup> However, acetylated cohesin, which enjoys a longer residence time, appears to accumulate at downstream regions of genes.<sup>295</sup> Analogously, condensin exhibits a bimodal distribution at transcription start and end sites<sup>296</sup>, and appears to translocate from loading sites in a similar transcription-dependent fashion, albeit with reduced processivity.<sup>282</sup> The distribution of SMC5/6 also appears to be similar, although uniquely, the complex is particularly highly associated with chromosome arms around DNA replication and loading is quantitatively correlated with chromosome length.<sup>283</sup> It appears therefore that translocation is common to the SMC complexes, and it seems to represent an effective mechanism to implement their structural functions without impeding other chromosomal activities.<sup>245</sup>

It is clear that further investigation is critical to improve understanding of these mechanisms of action, making the SMC complexes compelling targets for tool inhibitor development. Considering the data, it is reasonable to expect that a small molecule that interacts with either the hinge or head domains will impact complex function and be of significant scientific interest.

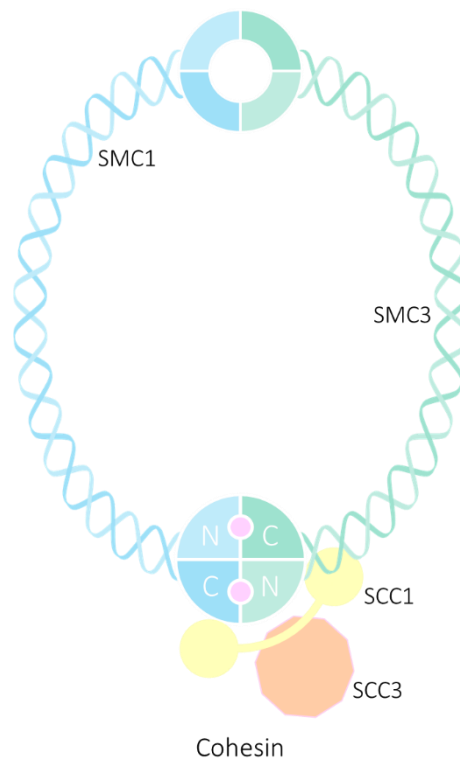


**Figure 3.3. Model for DNA Exit from and Entry to the Cohesin Ring.** (a) For DNA exit, lysine residues on the SMC3 head sense the DNA and trigger ATP hydrolysis to promote disengagement of the ATPase head, allowing DNA to pass through. ATP is then replenished, the head interaction is re-engaged and the kleisin dissociates from SMC3, allowing the DNA to exit the ring. (b) DNA entry occurs by a similar route but because the lysines point into the ring, this requires the cohesin complex to fold around the DNA, so as to expose these DNA-sensing residues. Figure adapted from Uhlmann, 2016.<sup>245</sup>

### 3.1.1 Cohesin

Cohesin is composed of two core SMC proteins, SMC1 and SMC3 and two non-SMC proteins, RAD21/SCC1 (sister chromatid cohesion protein 1) and SCC3<sup>251</sup>, Figure 3.4. Vertebrates have two SCC3 paralogues, stromal antigen 1 (SA1) and stromal antigen 2 (SA2). Cohesin-SA1 is specifically involved in telomere cohesion, whereas cohesin-SA2 is responsible for cohesion at centromeres.<sup>297</sup> SCC1 is a member of the kleisin family<sup>298</sup>, which, as discussed, interacts with the head domains of SMC3 and SMC1 via its N- and C-termini respectively to form a tripartite ring.<sup>222</sup> The HEAT repeat protein SCC3 interacts with SCC1, serving to further strengthen the cohesin ring structure.<sup>251</sup> It is understood that cohesin topologically embraces DNA within this ring.<sup>266</sup>

Although the primary function of cohesin is in regulating sister-chromatid cohesion, during the G1 phase it also engages in dynamic interactions between DNA binding sites, assisting in defining the spatial conformation of specific loci.<sup>299</sup> The complex is loaded on to chromatin by the SCC2-4 cohesin loader. Recruitment is likely via the SCC4 subunit, which serves as the chromatin adapter, to localise the loader to the DNA, while the SCC2 contacts cohesin.<sup>300</sup> Cohesin unloading is dependent on Wapl (wings apart-like protein homologue), which dissociates the SCC1 N-terminus from SMC3, thereby opening the kleisin gate.<sup>273</sup> In the absence of Wapl, cohesin is turned over more slowly, causing chromosomes to compact into wormlike 'vermicelli'.<sup>258</sup>



**Figure 3.4. Structure of the Cohesin Complex.** Structure comprised of core SMC proteins, SMC1 and SMC3 and two non-SMC proteins, kleisin SCC1 and HEAT repeat protein SCC3. Figure adapted from Losada & Hirano, 2005.<sup>250</sup>

Post-replication cohesin becomes cohesive, topologically embracing sister chromatids within its ring to maintain their cohesion.<sup>266</sup> This enduring sister chromatid cohesion is established during S phase by acetylation of conserved lysine residues on the SMC3 ATPase head by ECO1 (establishment of cohesion protein 1).<sup>301</sup> It appears that this acetylation locks the cohesin ring by preventing DNA from triggering ATP hydrolysis and by recruiting sororin to inhibit the complex from interacting with Wapl, which in combination would complete the dynamic binding and dissociation cycle.<sup>302,303</sup> The obvious predicament here is that successful DNA loading requires active DNA-sensing lysines. As a result, acetylation must be postponed until both chromatids are adequately associated with the cohesin complex.

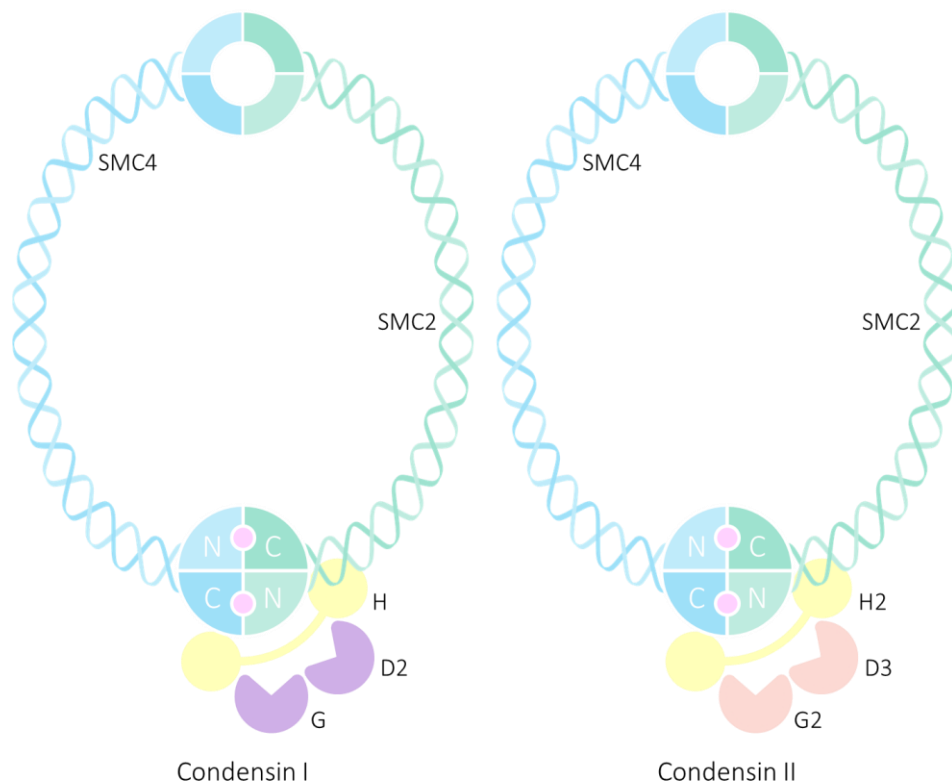
There are two plausible mechanisms that could enable co-entrapping of two sister DNA strands by cohesin. One possibility is that replication forks can pass through cohesin rings. In this case, acetylation at any time prior to or during fork passage would establish sister-chromatid cohesion. The second is that acetyltransferases follow immediately behind replication forks, such that as cohesin establishes links between the sister chromatids, which are naturally in close proximity, these are stabilised.<sup>245</sup> It is likely that both of these mechanisms contribute to establishing cohesion. The two human ECO1 paralogues, ESCO1 (establishment of cohesion 1 homologue 1) and ESCO2 (establishment of cohesion 1 homologue 2), contribute non-redundantly to sister chromatid cohesion. While ESCO2 interacts with replication machinery and therefore likely follows behind a replication fork<sup>304</sup>, ESCO1 acetylates cohesin as early G1, independently of DNA replication.<sup>305</sup>

As cells enter mitosis, cohesin must be removed from the chromosomes to enable proper chromosome segregation. By necessity this dissociation is tightly regulated and occurs during two phases of mitosis.<sup>306</sup> The majority of cohesin from the chromosome arms is removed in the 'prophase pathway'. Mitotic kinases phosphorylate sororin, to destabilise its interaction with cohesin, enabling the complex to be unloaded by Wapl.<sup>307,308</sup> The sororin at the centromeres however is protected from phosphorylation throughout prophase by shugoshin. Cohesion is therefore maintained here until SCC1 is cleaved at the metaphase-anaphase transition by a protease named separase.<sup>308,309</sup>

Cohesin also has additional cellular roles in DNA damage response. SCC1 was in fact first identified because its mutation in *S. pombe* resulted in cells which were hypersensitive to UV and IR radiation.<sup>260</sup> It is understood that cohesin has a specific and vital role in repair of DNA double stranded breaks via sister chromatid homologous recombination during S and G2 phases of the cell cycle.<sup>310</sup> HR-mediated DSB repair requires an undamaged DNA template. During the mitotic cell cycle, use of the sister chromatid is preferred in order to prevent potential loss of heterozygosity. It is likely that sister-chromatid cohesion near to the DSB, which keeps the undamaged chromatid in close proximity, should promote strand invasion and therefore DSB repair by HR.<sup>251</sup> The observed recruitment of cohesin to DSBs supports this suggested role in DNA repair.<sup>311</sup> Furthermore, a study in *S. pombe* indicated that use of non-cleavable SCC1 or inactivation of separase hindered DNA repair at G2, indicating a necessary requirement for cohesin cleavage by separase in DNA repair.<sup>312</sup> Further studies have implicated cohesin in DNA damage checkpoint activation at S-phase<sup>313</sup> and more recently at G2/M.<sup>314</sup>

### 3.1.2 Condensin

In vertebrates there are two different condensin complexes, condensins I and II. These complexes are composed of five subunits: the core SMC proteins, SMC2 and SMC4, which are common to both, and three chromosome-associated proteins (CAPs). Condensin I incorporates CAPD2, CAPH and CAPG, while condensin II contains CAPD3, CAPH2 and CAPG2<sup>250</sup>, Figure 3.5. CAPH and CAPH2 are members of the aforementioned kleisin family<sup>298</sup>, while CAPD2, CAPD3, CAPG and CAPG2 contain HEAT repeats, which serve as flexible scaffolding to afford secure protein-protein interactions<sup>315</sup> and are suggested to elicit DNA binding activity.<sup>316</sup> Available data suggests that, like SCC1, the kleisins bind across the SMC subunits. The N-terminal half of CAPH/H2 links CAPD2/D3 to SMC2 and the C-terminal half links CAPG/G2 to SMC4. There is no direct interaction between the HEAT subunits and SMC proteins.<sup>317,318</sup> Unlike in cohesin, where the SMC coiled-coil arms are spread apart, in condensin they often appear closer together such that with the regulatory subunits bound at the head domain the complex resembles a tadpole.<sup>265,275</sup> However, it seems this rod-like structure opens to a ring upon DNA binding.<sup>319</sup>



**Figure 3.5. Structures of the Condensin Complexes.** Structures comprised of core SMC proteins, SMC2 and SMC4 and three non-SMC proteins, kleisins CAPH or CAPH2 and HEAT repeat proteins CAPD2 or CAPD3 and CAPG or CAPG2. Figure adapted from Losada & Hirano, 2005.<sup>250</sup>

The primary role of condensins is in chromosome condensation during mitosis. Although both play essential roles in mitotic chromosome architecture and segregation, condensins I and II display different behaviours during the cell cycle<sup>320</sup> and different roles in condensation.

Condensin II is present on DNA throughout interphase and concentrates at prophase chromosomes, suggesting that it participates in early stage chromosomal condensation. Condensin I however only concentrates on chromosomes after nuclear envelope breakdown (NEBD) at prometaphase.<sup>321</sup> Until recently it was believed that condensin I is segregated to the cytoplasm and therefore only gains access to chromosomes after NEBD, however it is now understood that a small pool of condensin I persists in the nucleus during G1, which is gradually lost during S and G2.<sup>322</sup> During prometaphase the condensins work in tandem to afford well resolved sister chromatids by metaphase, which enables them to be successfully separated at anaphase.<sup>323</sup> Condensin I appears to be responsible for lateral compaction and in its absence the mitotic chromosomes become wider and shorter, whereas if cohesin II is removed, a loss of axial condensation is evident with longer and less rigid chromosomes.<sup>324</sup>

The interaction of condensin with DNA is less well understood than that of cohesin. In 2011 Cuylen *et al.* showed that circular minichromosomes linked to condensin disengaged upon either proteolytic opening of the condensin or linearization of the DNA, suggesting that condensin topologically entraps DNA.<sup>267</sup> There is accumulating evidence that ATPase activity is essential for condensin binding to DNA. ATP binding mutations at the Walker A or Walker B motifs of SMC2 or SMC4 prevent condensin loading, while ATP hydrolysis mutations permit loading but disrupt condensin localisation and higher order structure.<sup>325,326</sup> There is also data to suggest that the HEAT subunits<sup>316</sup> and the SMC hinge domains<sup>327</sup> influence DNA binding.

The molecular mechanism of action of the condensins is hotly debated. Condensin has two well-described enzymatic functions: ATPase-dependent positive supercoiling of dsDNA and ATP-independent DNA reannealing.<sup>321</sup> It has been shown that condensin is able to introduce positive supercoils in DNA.<sup>328</sup> Such large-scale folding of chromatin intensifies torsional stress in the DNA<sup>329</sup>, which must be relieved periodically to allow folding to continue. Topoisomerases such as Top2 (topoisomerase II) are therefore essential for hypercompaction of chromosomes.<sup>330</sup> Top2 also serves to decatenate DNA as it is condensed<sup>331</sup>, to facilitate individualisation and sister chromatid resolution. The reannealing activity of condensin is necessary to eliminate unwound DNA generated by transcription, which is believed to be essential for successful supercoiling.<sup>296</sup> As a result, classically, an integrated model is speculated<sup>323</sup>, in which condensin generates dsDNA by inducing reannealing<sup>276</sup> and introduces supercoils to cause superhelical tension<sup>332</sup>, thereby promoting Top2-mediated decatenation<sup>333</sup> and chromosome resolution. Continued positive supercoiling could then cause the formation of chiral loops, which could subsequently be organised into a solenoidal structure by higher order assembly.<sup>334</sup> In support of this model, in cross sections of mitotic chromosomes, strands comprised of condensin, topoisomerase II and



KIF4A (kinesin family member 4A) are visible in the central core.<sup>335</sup> These strands could represent the central scaffold structure associated with higher order assembly.

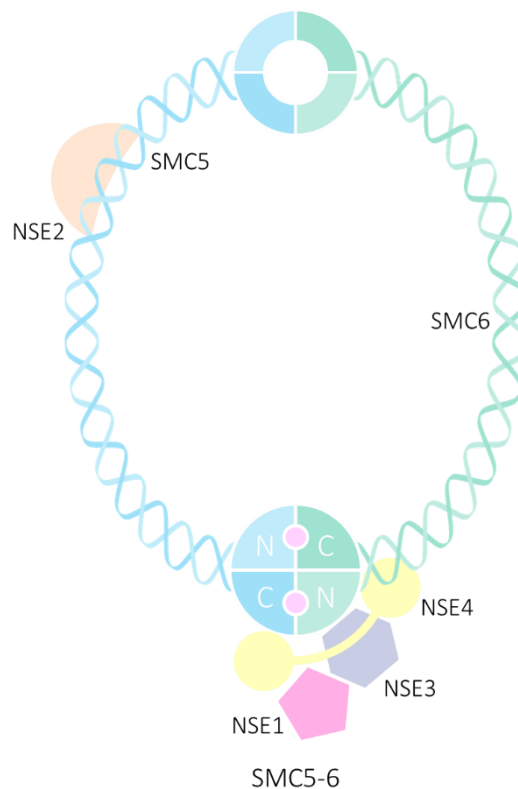
The lack of a molecular explanation for how such a central scaffold would form, has led to the proposal of alternative models for the condensin mechanism of action. One such model is built on the observation of condensin at the intersections of 10 nm fibres.<sup>336</sup> This has resulted in the proposal that condensin functions by stochastically crosslinking freely diffusing chromatin fibres.<sup>245</sup> *Ab initio* simulations of a chromosome constrained by stochastic interactions between condensin-binding sites, successfully generate behaviour closely resembling that observed *in vivo*.<sup>337</sup> Another popular model for condensin action is the loop extrusion model. This suggests that condensin binds at two nearby points on DNA and slides in such a way as to gradually generate large loops. Polymer simulations of chromosome dynamics have been used to show that condensin could bind to two adjacent loci and slide the contacts in opposite directions to create such extruded loops.<sup>338</sup>

Like cohesin, condensin has additional roles in DNA damage response. In 2002 an *S. pombe* temperature-sensitive *cnd2* (orthologous to human CAPH) mutant was isolated. As well as the expected mitotic chromosome condensation defects, the mutant displayed increased sensitivity to UV, hydroxyurea and methyl methanesulfonate and a defect in activation of CHK2, a checkpoint kinase which is activated in response to DNA damage.<sup>259</sup> These results suggest a role for condensin in replication checkpoint control and DNA repair. Since this, the human condensin complexes have also been implicated in DNA repair. Condensin I appears to be involved in repair of DNA single-stranded breaks in association with PARP1 and XRCC1<sup>339</sup>, whereas condensin II is thought to have a role in HR repair of DSBs in association with MCPH1 (microcephalin 1).<sup>340</sup> In addition to being directly involved in DNA repair to help maintain genome integrity, it has been found that in yeast condensin prevents unwanted intrachromosomal HR at the rDNA locus and controls rDNA stability.<sup>341</sup> Such dysregulation of ribosome biogenesis has been linked to cancers and other human diseases.<sup>251</sup>

### 3.1.3 SMC5/6

The human SMC5/6 complex comprises of two core SMC proteins SMC5 and SMC6, as the name suggests, and four non-SMC elements, NSE1, 2, 3 and 4, Figure 3.6.<sup>342</sup> SMC5 and SMC6 are substantially divergent from SMC1-4. NSE3 and NSE4 independently bind at the head domains of SMC5 and SMC6. NSE4 is another member of the kleisin protein family, which binds across the SMC head domains.<sup>277</sup> NSE3, also known as MAGEG1, belongs to the MAGE (melanoma-associated antigen) protein class, which are overexpressed in certain cancers. MAGE proteins

have suggested functions in cell cycle regulation, apoptosis and neuronal development.<sup>343</sup> NSE3 interacts with NSE1 which contains a RING (really interesting new gene) finger domain that confers ubiquitin ligase activity.<sup>344</sup> Interestingly, NSE1 and NSE3 are similar in structure to the prokaryotic SMC subunits ScpB and MufE, which may suggest functional similarities between SMC5/6 and the prokaryotic SMC complex.<sup>345</sup> Finally, uniquely, SMC5/6 incorporates a small ubiquitin-like modifier (SUMO) ligase, NSE2. NSE2 binds to the SMC5/6 dimer independently of NSE1, 3 and 4, through interaction with the SMC5 coiled-coil.<sup>346</sup> It promotes the covalent attachment of SUMO to proteins. This SUMOylation activity is required in DNA damage response, playing a key role in maintaining genomic stability.<sup>347</sup> While NSE2 is essential, its SUMO ligase activity does not appear to be.<sup>226</sup> In humans, NSE2 does not affect the stability of the SMC5/6 complex, but the SUMO-ligase activity is executed only within the context of the complex.<sup>343</sup>



**Figure 3.6. Structure of the SMC5/6 Complex.** Structure comprised of core SMC proteins, SMC5 and SMC6 and four non-SMC proteins, kleisin NSE4, SUMO ligase NSE2 and ubiquitin ligase NSE1-NSE3.

The function of the SMC5/6 complex is less well understood than that of cohesin and condensin. It is known to play a prominent role in DNA repair by recombination. SMC6 was first identified via a screen for radiation-sensitive mutations in *S. pombe*<sup>348</sup> and further studies have shown that mutation of any of the subunits of the SMC5/6 complex induces hypersensitivity to DNA damage agents.<sup>343,347,349</sup> In particular, SMC5/6 is found to be recruited to DSBs where it promotes repair

by sister-chromatid HR, possibly in conjunction with cohesin.<sup>255</sup> In addition, the complex has been shown to be involved in repair of collapsed replication forks, such that inactivation causes accumulation of X-shaped HR intermediates<sup>254</sup>, and resolution of sister-chromatid linkages during mitosis.<sup>350</sup> These investigations indicate that the SMC5/6 complex promotes HR-dependent rescue of stalled replication forks.<sup>251</sup> Like condensin, the SMC5/6 complex also has a suggested role in maintaining rDNA stability by helping to prevent unwanted HR at the rDNA locus, thereby suppressing recombinational loss of rDNA repeats.<sup>351</sup>

It is unlikely though that DNA repair is the essential function of SMC5/6, as other recombination proteins are dispensable for eukaryotic survival.<sup>352</sup> The search for the true essential function of the complex is ongoing. A study aimed at examining the function of SMC5/6 throughout the cell cycle determined that the complex associates with chromatin during interphase but largely dissociates at mitosis. Depletion of SMC5/6 was observed to cause deformation of sister chromatids, abnormal Top2 distribution to distal arms rather than centromeres, and defective chromosomal segregation. Interestingly, this damage can be substantially alleviated by allowing the cells extra time to complete DNA replication before mitosis by subjecting them to G2 arrest. This could suggest that SMC5/6 depletion either slows DNA replication or provokes a temporal disconnection between replication and chromosome assembly, possibly due to defective checkpoint control.<sup>353</sup> Congruous with these findings, another recent study revealed that the essential functions of SMC5/6 manifest selectively in G2/M. Critical roles of SMC5/6 were identified in facilitating replication past replication fork barriers (RFBs), and metabolising DNA damage tolerance (DDT) intermediates, formed in response to endogenous replication stress, such as the RFBs. This suggests that SMC5/6 expedites replication whilst helping to regulate genome integrity.<sup>354</sup>

Others have also noted chromosome segregation errors resulting from SMC5/6 depletion. Cohesin is the main constituent of chromatid cohesion. However, Farcas *et al.* have shown that cohesin inhibits Top2 from resolving sister chromatid intertwinings (SCIs), suggesting that these could also contribute to cohesion.<sup>355</sup> SCIs arise behind replication forks when the advancing fork occasionally rotates with the turn of the DNA helix. This leaves the sister chromatids entwined and if left unresolved will prevent chromosome segregation during mitosis. In 2014, Jeppsson *et al.* showed that cohesin and cohesion motivate the chromosomal association and localization of SMC5/6, which interacts at SCIs and facilitates their Top2-independent resolution.<sup>283</sup>

Other novel roles of the complex are still being discovered. A genome-wide analysis of the replication dynamics of a *S. cerevisiae* SMC6 mutant has recently been conducted.<sup>356</sup> The study highlighted roles of SMC5/6, firstly in preventing misincorporation of ribonucleoside

monophosphates (rNMPs) into DNA, and secondly ensuring proper DNA repair after incision to remove any rNMPs incorporated erroneously. The mutant was observed to incorporate ribonucleotides and accumulate ssDNA gaps, capable of producing toxic HR intermediates.<sup>356</sup>

How SMC5/6 interacts with DNA and chromosomes also remains elusive. In 2011, Roy *et al.* reported that SMC5 and SMC6 preferentially bind ssDNA<sup>357,358</sup>, consistent with their roles in DNA repair. Concordantly, last year it was demonstrated that the SMC5/6 hinge has significantly higher affinity for ssDNA than dsDNA.<sup>223</sup> In this publication, Alt *et al.* determined the X-ray crystal structure of the *S. pombe* SMC5/6 hinge. They identified two features, a 'molecular latch' and a 'hub', which are critical for proper SMC5/6 function. The 'latch' comprises of 13 amino acids, 607-619, of SMC5, which compose an extended  $\beta$ -hairpin that makes several interactions with the SMC6 hinge. The 'hub' is the cluster of interacting SMC6 residues, residing at the junction of the SMC6 arm and globular hinge domain. They speculate that the distinctive structure of SMC5/6 facilitates dynamic loading and unloading. Interestingly, it was not possible to fit the experimental data to a 1:1 binding model. The authors instead suggest that the association and dissociation of SMC5/6 is biphasic, with ssDNA making a primary interaction with SMC5-R609/R615 at the latch, and a secondary interaction with residues at the hub.<sup>223</sup>

While it seems SMC5/6 preferentially binds ssDNA, it is also known that the ATPase activity of the complex is stimulated chiefly by dsDNA and only to a lesser extent ssDNA.<sup>359</sup> In 2015, Kanno *et al.* were able to show that purified SMC5/6 binds dsDNA via a mechanism dependent on ATP hydrolysis.<sup>268</sup> They observed that the association of SMC5/6 with circular, but not linear, DNA was resistant to high salt concentrations, suggestive of a topological interaction. Additionally, they established that the interaction promotes Top2 dependent catenation, indicating that the complex can tether two DNA molecules. They therefore suggest that SMC5/6 acts to interconnect DNA molecules via ATP-regulated topological entrapment, potentially actually inducing SCIs behind the fork to facilitate fork rotation and expedite replication.<sup>268</sup> In 2016, Zabradý *et al.* described a novel DNA-binding mode of SMC5/6 via its NSE1/3/4 subcomplex.<sup>360</sup> They determined that the NSE3 protein binds to dsDNA via its evolutionarily conserved C-terminal winged-helix domain. NSE1 and NSE4 may also interact with the DNA. Disruption of the DNA-binding ability of NSE3 is lethal in *S. pombe*, analogous to gene deletion.<sup>360</sup> In the knowledge that SMC5 and SMC6 preferentially bind ssDNA<sup>357,358</sup>, but ATP hydrolysis is provoked by dsDNA<sup>359</sup>, the authors suggest that the NSE subunits interact with dsDNA and subsequently stimulate ATPase activity at the SMC heads.<sup>360</sup>

If any SMC5, SMC6 or NSE mutations prove non-lethal, the SMC5/6 complex may also become a plausible target for design of a novel cancer therapeutic utilising a synthetic lethality approach.

As discussed in Chapter 1.3, such an approach is reliant on genetic interactions of two mutations for which neither mutation alone will kill cells, but that in combination render them inviable. This could create an opportunity to selectively kill cancer cells that harbour mutations in genes synthetic lethal with SMC5/6 by mimicking the effect of the SMC5/6 mutation using targeted drug therapy.

It is also established that SMC5/6 has a role in the alternative lengthening of telomeres (ALT) pathway.<sup>361</sup> Telomeres are proteinaceous, repetitive DNA elements, attached at the end of each chromosome<sup>362</sup>, which are shortened at every cell division owing to the end-replication problem of the lagging strand.<sup>363</sup> Ultimately, shortened telomeres result in cell senescence and as such this effectively limits the proliferative potential of the cell.<sup>364</sup> As discussed in Chapter 1.2, by upregulating telomerase transcription, which is repressed in normal somatic cells, cancer cells can overcome this limitation, affording them replicative mortality.<sup>365</sup> 10-15% of cancer cells however rely on the ALT pathway instead of telomerase reactivation.<sup>366</sup> Here HR between telomeric sequences is used to elongate telomeres.<sup>367</sup> One defining feature of ALT cells is that telomeres become localised to promyelocytic leukaemia (PML) bodies, named ALT-associated PML bodies (APBs).<sup>368</sup> It has been shown that SMC5/6 is required for telomere recombination and APB formation. In particular NSE2 appears to be responsible for SUMOylation of several shelterin subunits.<sup>361</sup> The shelterin complex is responsible for protecting telomeres from being recognised as DSBs and therefore preventing telomere HR.<sup>369</sup> SUMOylation often changes protein-protein interactions and it is possible that the stability of the complex could be affected and disassemble to leave the telomeres unprotected and exposed for HR. In support of this model, Chavez *et al.* report similar observations in *S. cerevisiae*.<sup>370</sup> Typically, telomerase is constitutively expressed in *S. cerevisiae* and if it is genetically inactivated this causes telomere loss and senescence. However, some telomerase-null yeast are able to survive, utilising HR-dependent telomere maintenance to slow senescence, echoing ALT cancer cells. Furthermore, SUMOylation-deficient telomerase-null cells senesce at an elevated rate and NSE2 is identified as the key ligase mediating this effect.<sup>370</sup> Whatever the mechanism, there is good evidence that depletion of SMC5/6 subunits causes inhibition of telomere HR, and should thereby induce telomere shortening and cell senescence in the tumour cells. Interestingly, in 2016 it was discovered that SMC5/6 is present at telomeres throughout the cell cycle and this association is dependent on NSE3.<sup>371</sup> Cells mutated in NSE3 are defective in SMC5/6 localisation and have shortened telomeres. This is consistent with the prior observation that SMC5/6 interacts with DNA via the NSE3 subunit,<sup>360</sup> and suggests an additional role for SMC5/6 in telomere maintenance, separate from the established role in HR-mediated telomere elongation.

## 3.2 Results

### 3.2.1 SMC5 and SMC6

DELTA-BLAST<sup>180,183</sup> searches for human homologues of SMC5 and SMC6 identified the other SMC proteins and RAD50, Table 3.1 and Table 3.2. The Max and Total scores are associated with the statistics of the database search. The Max score depends on the query-hit alignment length and quality of the ‘best’ region of similarity. The Total score is equal to the Max score if only a single alignment is present. When multiple aligned regions exist, each individual score contributes to the Total score, and therefore this is greater than the Max, as was often observed. The Query cover gives the length of the alignment as a percentage of the length of the query sequence. In this work, all hits showed Query cover greater than 90%, demonstrating good conservation. The E (expectation) value evaluates the likelihood that the hit is significant by calculating the number of hits of this calibre expected due to chance alone. This depends on the quality of the alignment and the size of the database. A number close to zero indicates a significant hit. For the SMC proteins there was high confidence that the hits were not due to chance. Finally, the Ident represents the percentage identity between the query and hit by simple sequence alignment.<sup>372</sup> Despite the Query cover showing that the alignments spanned the majority of the length of the query sequence, the Ident indicated that the number of identical residues by sequence alignment was low. This could predict difficulties with modelling on distant relatives.

**Table 3.1. SMC5 Human Homologues.** Results of DELTA-BLAST<sup>180,183</sup> search on protein sequence.

Description	Max Score	Total Score	Query Cover	E Value	Ident
SMC5	629	629	100%	0.E+00	100%
SMC6	228	228	96%	3.E-61	15%
SMC2	204	204	92%	2.E-53	12%
SMC3	188	188	92%	2.E-48	13%
SMC4	180	316	92%	5.E-46	12%
SMC1b	131	260	92%	1.E-30	10%
SMC1a	130	252	92%	1.E-30	12%
RAD50	94	238	91%	3.E-19	10%

EMBL-EBI Clustal Omega<sup>181,182</sup> was used to compile a multiple amino acid sequence alignment of SMC5, SMC6 and the identified human homologues, Figure Apx.1. The alignment indicated relatively good conservation at the ATPase head domain but poor conservation at the hinge.

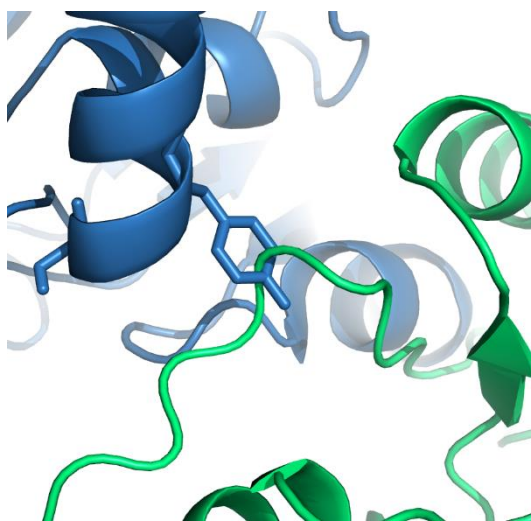
A DELTA-BLAST<sup>180,183</sup> search was used to identify relevant structural data available in the PDB ([www.rcsb.org](http://www.rcsb.org)).<sup>158</sup> Although there were no structures of human SMC5 and SMC6, potential model templates were available. These templates were separated into two categories, exhibiting sequence similarity with either the hinge or head domains.

**Table 3.2. SMC6 Human Homologues.** Results of DELTA-BLAST<sup>180,183</sup> search on protein sequence.

Description	Max Score	Total Score	Query Cover	E Value	Ident
SMC6	655	655	100%	0.E+00	100%
SMC4	235	235	94%	2.E-63	12%
SMC2	231	231	94%	3.E-62	14%
SMC5	219	219	93%	1.E-58	14%
SMC3	194	194	93%	3.E-50	12%
SMC1a	153	348	93%	1.E-37	12%
SMC1b	142	280	93%	5.E-34	12%
RAD50	114	195	92%	1.E-25	12%

#### *Computational Assessment of SMC5/6 Hinge Domain*

Despite the multiple amino acid sequence alignment<sup>181,182</sup> of SMC5, SMC6 and human homologues indicating poor conservation at the hinge domain and the only available template PDB structures being non-human paralogues an attempt was made to model the SMC5/6 hinge. The SMC5/6 domain was modelled on the X-ray crystal structure of the hinge domain of the *Thermotoga maritima* SMC protein (PDB ID: 1GXJ)<sup>222</sup> using SWISS-MODEL.<sup>179</sup> This was unsuccessful. SMC5 and SMC6 had to be modelled independently on 1GXJ chains A and B respectively, which resulted in regions clashing at the interface, Figure 3.7.



**Figure 3.7. Model of Human SMC5/6 Hinge Domain by SWISS-MODEL.**<sup>179</sup> SMC5 shown in blue and SMC6 in green, illustrating clashing at interface.

In contrast, the multiple amino acid sequence alignments of SMC5 and SMC6 with their respective orthologues indicated good conservation at the hinge domain, Figure Apx.2 and Figure Apx.3. Most notable were the SMC5 Ser-624 and Tyr-626 residues. Recently, in *S. pombe* it was observed that SMC5 mutations at Ser-610 and Tyr-612 disrupt the SMC5/6 hinge interaction.<sup>223</sup> In this work, these serine and tyrosine residues were found to be conserved in humans, as well as all other orthologues compared, Figure 3.8.

SMC5\_YEAST (623-632) - I K Q S A Y G S K Q - *Saccharomyces cerevisiae*  
 SMC5\_SCHPO (607-616) - V F R S A Y G D R E - *Schizosaccharomyces pombe*  
 SMC5\_TAKRU (612-621) - L K R S V Y S K M I - *Takifugu rubripes*  
 SMC5\_CHICK (600-609) - I K V S T Y T K L S - *Gallus gallus*  
 SMC5\_XENLA (601-610) - T K K S V Y S Q K L - *Xenopus laevis*  
 SMC5\_HUMAN (621-630) - V K T S F Y S N K V - *Homo sapiens*  
 SMC5\_MOUSE (621-630) - L K T S V Y S N K V - *Mus musculus*

**Figure 3.8. Multiple Sequence Alignment of SMC5 Orthologues.** Conserved Ser and Tyr highlighted in orange.

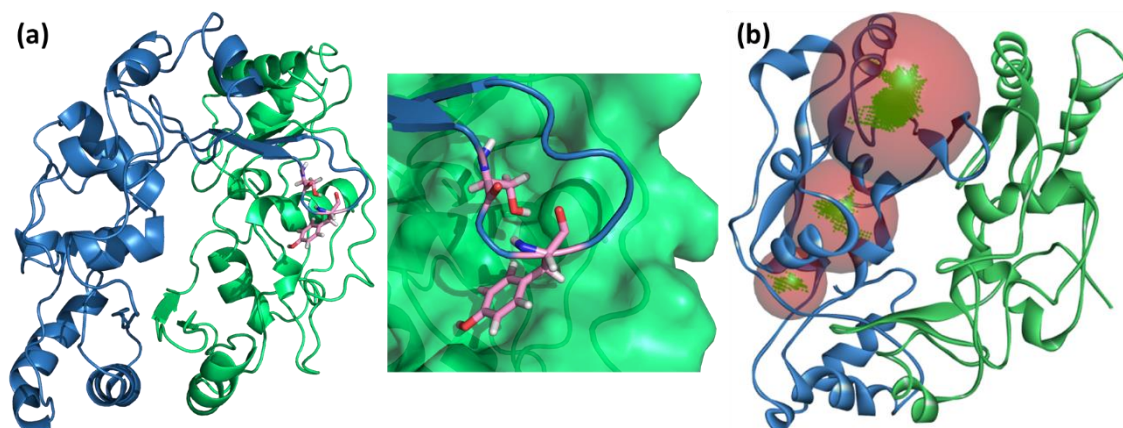
Given the improved homology of the SMC5/6 orthologues compared to the paralogues, these represented more desirable template structures. A crystal structure of the *S. pombe* SMC5/6 hinge domain (PDB ID: 5MG8) was identified.<sup>223</sup> The SMC5/6 hinge was modelled on this structure using MODELLER 9.12<sup>177</sup>, which unlike SWISS-MODEL<sup>179</sup>, permitted the two chains to be modelled simultaneously. The resulting model was prepared and minimised using Discovery Studio 4.0<sup>186</sup>, Figure 3.9 (a). The SMC5 Ser-624 and Tyr-626 residues are positioned at the interface with SMC6 suggesting that mutating these residues in humans would likely cause the same disruption of the SMC5/6 hinge as was observed in *S. pombe*.<sup>223</sup>

One criteria for a good drug target is the ability to develop an assay against the protein. This is important in enabling screening and drug development. A drug that is designed to mimic the SMC5 Ser-624 and Tyr-626, identified as key residues in stabilising the SMC5/6 hinge interaction, would target SMC6. Unfortunately, SMC6 had proved unstable, suggesting that assay development would be challenging (Alt, A., Genome Damage and Stability Centre, University of Sussex, personal communication). As a result, SMC5 was the preferred target. Accordingly, Discovery Studio 4.0<sup>186</sup> was used to identify potential binding pockets from receptor cavities on SMC5. It was hoped that structural pockets would be identified at the SMC5-SMC6 interface, which might indicate the potential to target the protein using an SMC6 mimetic, e.g. an  $\alpha$ -helical mimetic, although, by eye, the protein-protein interaction (PPI) interfaces looked to be composed only of  $\beta$ -sheets, which are notoriously difficult to target. Unfortunately, no structural pockets were identified at the PPI interfaces by Discovery Studio 4.0<sup>186</sup>. A subset of the structural pockets identified are shown in Figure 3.9 (b).

At the time this work was completed, the SMC5 hinge domain therefore appeared undruggable. The interface is composed only of  $\beta$ -sheets with no discernible structural pockets, affording no scope for rational design of a PPI inhibitor. No further structures existed of other proteins or small molecules bound at the hinge, which could provide inspiration for other druggable sites or the corresponding pharmacophores. In addition, it was assumed that a fragment screening approach would be unsuccessful. Due to their small size, the majority of fragments exhibit low affinity. It is therefore essential that any screening is carried out against unliganded protein with the binding site accessible.<sup>373</sup> This would be problematic, as SMC5 is known to be less stable in



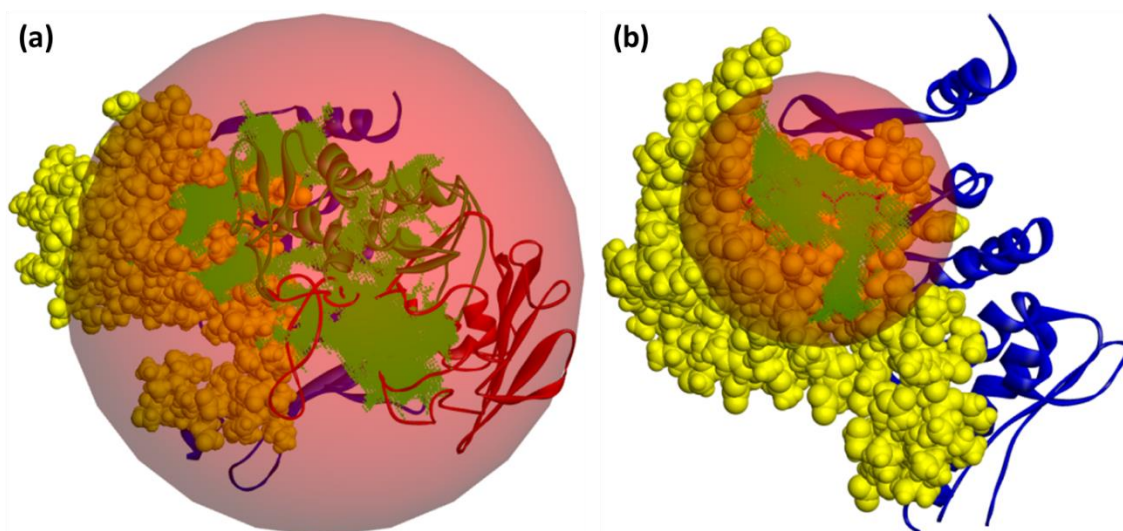
the absence of SMC6, which will hinder assay development (Alt, A., Genome Damage and Stability Centre, University of Sussex, personal communication).



**Figure 3.9. Model of Human SMC5/6 Hinge.** Human SMC5/6 hinge modelled on 5MG8.<sup>223</sup> SMC5 shown in blue and SMC6 in green. (a) Prominent Ser-624 and Tyr-626 residues highlighted in pink. (b) Structural pockets highlighted in image from Discovery studio 4.0.<sup>186</sup>

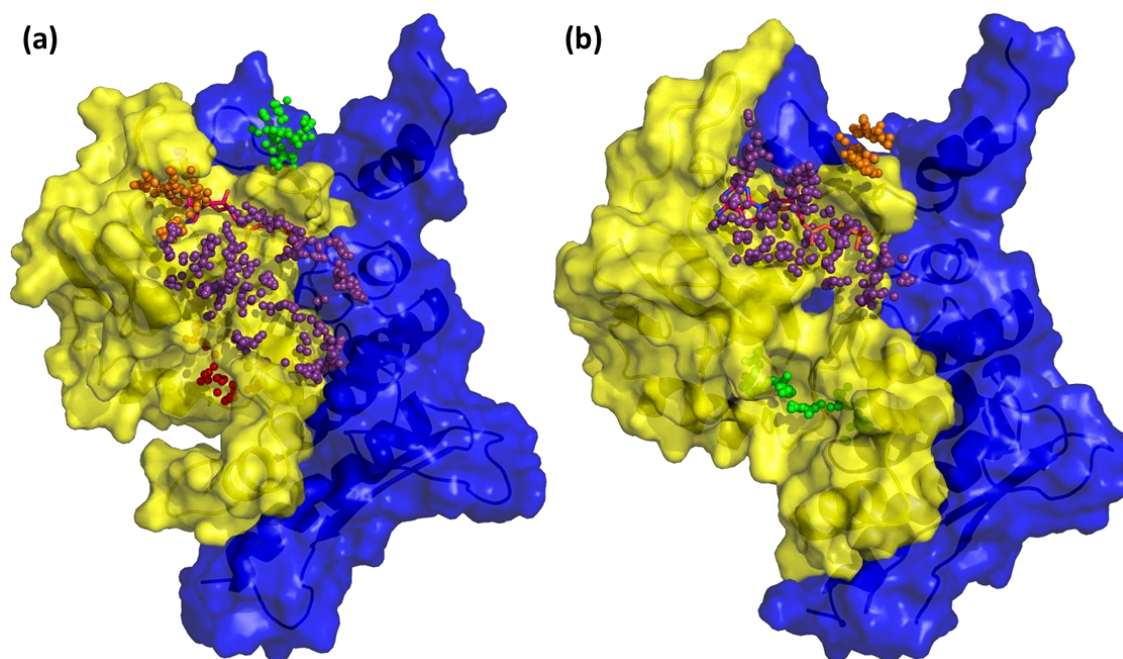
#### *Computational Assessment of SMC5/6 Head Domain*

Unlike at the hinge, the multiple amino acid sequence alignment<sup>181,182</sup> of SMC5, SMC6 and other identified human homologues indicated good conservation at the head domain, Figure Apx.1, suggesting that a good model could be constructed even from a relatively distant relation. In addition, the DELTA-BLAST<sup>180,183</sup> search identified more reported structures of better sequence similarity to the head domains than to the hinge. Many of these crystal structures were of little use, describing the structure of the SMC head domain at either the N- or C-terminal regions but not the complete structure. Nevertheless, the most suitable available crystal structure was identified and the SMC5/6 head was modelled on the *Pyrococcus furiosus* SMC ATPase (PDB ID: 1XEX)<sup>224</sup>, using MODELLER 9.12.<sup>177</sup>



**Figure 3.10. SMC5/6 Head with Structural Pockets Identified.** SMC5 N- and C-terminal chains shown in yellow and blue respectively and SMC6 in red and green. (a) Model of human SMC5/6 head with first structural pocket identified. (b) 1XEX<sup>224</sup> SMC5 with structural pocket identified at ATP binding domain. Images produced in Discovery studio 4.0.<sup>186</sup>

Discovery Studio 4.0<sup>186</sup> was used to prepare and minimise the model, and to identify potential binding pockets from receptor cavities. The first pocket identified is illustrated in Figure 3.10 (a). This large pocket was splayed, extending throughout the SMC5/6 model, and appeared to comprise of a collection of smaller pockets. These included cavities at both ATP binding sites. With an aim to check that it was possible to correctly identify the shallow ATP binding pockets, the 1XEX<sup>224</sup> template was examined. Gratifyingly, the ATP binding domains were successfully identified. Figure 3.10 (b) shows the pocket identified at the SMC5 N-terminus, which encompasses the ATP binding site and a second branching pocket. It should be noted that, as was true of the hinge domain, the SMC5 head domain protein was more stable than the SMC6 head domain, and attempts to prepare and purify SMC6 head domain protein had been unsuccessful (Alt, A., Genome Damage and Stability Centre, University of Sussex, personal communication). Therefore, SMC5 was again the preferred target.



**Figure 3.11. Structural Pockets Identified by Fpocket 2.0.**<sup>165</sup> SMC N- and C-terminal chains shown in yellow and blue respectively with pockets highlighted and ATP from 1XEX template in pink. (a) Model of human SMC5 head. (b) 1XEX template.<sup>224</sup>

**Table 3.3. Druggability Prediction Scores from Fpocket 2.0.**<sup>170</sup>

	Pocket	Colour	Druggability Score
Human SMC5/6 Model	2	Purple	0.046
	10	Orange	0.014
	12	Green	0.012
	31	Red	0.109
1XEX Template	1	Purple	0.087
	7	Orange	0.016
	15	Green	0.252

With an aim to improve understanding of the available structural pockets at and around the SMC5 N-terminal ATP binding site, Fpocket 2.0<sup>165</sup> was employed. Fpocket 2.0<sup>165</sup> is an open source, cavity prediction algorithm, which is coupled to a structure-based target druggability prediction score.<sup>170</sup> 32 pockets were identified for the SMC5/6 model and 25 for the 1XEX template.<sup>224</sup> The pockets at and nearest to the SMC N-terminal ATP binding site are illustrated in Figure 3.11 and the associated Fpocket druggability scores<sup>170</sup> are listed in Table 3.3.

Typically, a druggable pocket should exhibit an Fpocket druggability score of higher than 0.5, which suggests that the ATP binding site is not druggable. However, the most important descriptor in calculating this druggability score is the mean local hydrophobic density of the binding site, with the hydrophobicity and normalised polarity scores, referring to the physicochemical character of the residues lining the pocket, also taken into account.<sup>170</sup> As such, solvent exposed cavities are often predicted as undruggable. In addition, ATP binding sites are designed to accommodate charged phosphate residues, requiring a polar environment, which is likely to negatively impact the druggability scores of neighbouring pockets. In summary, the druggability prediction scores generated may not be truly reflective of the druggability of the identified pockets. Furthermore, despite the poor druggability scores, Fpocket 2.0<sup>165</sup> and Discovery Studio 4.0<sup>186</sup> concurred in identifying the same adenosine binding site and the same second branching pocket. The decision was made to analyse the pockets further using computational and experimental approaches in attempt to determine the true viability of the SMC5 head domain as a target.

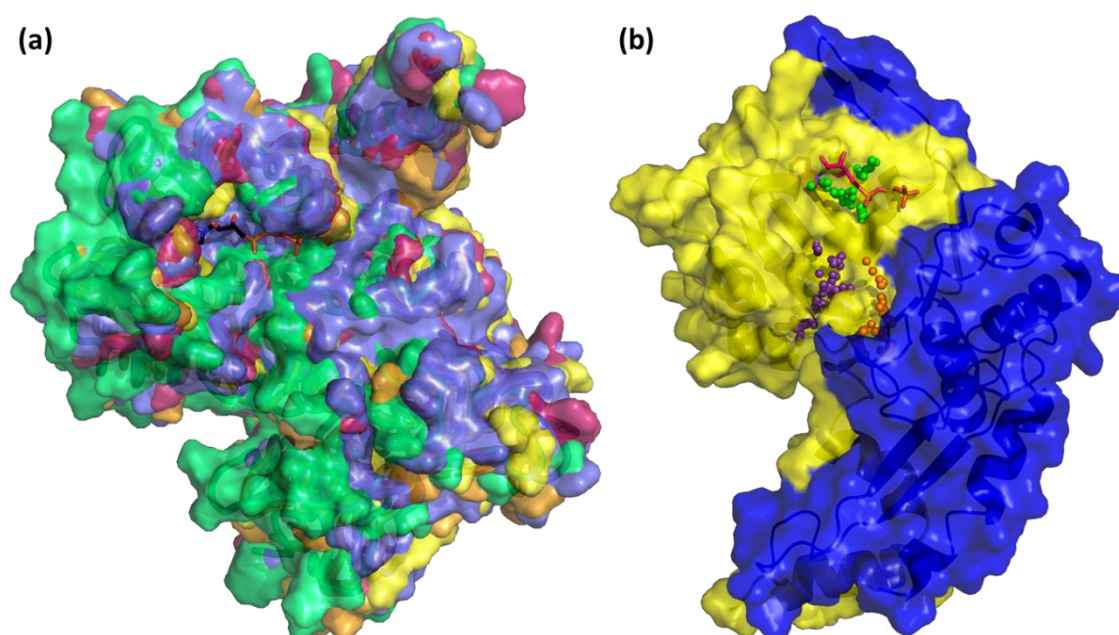
Having identified the adenosine binding pocket and branching channel as potentially druggable sites, further information was required to assess the quality of the target and likely specificity that could be achieved. Two questions were proposed:

1. Is there sufficient confidence that the pockets exist?
2. Are the pockets conserved in other proteins?

#### *Improving Confidence in the Pockets*

As a first attempt to improve confidence in the pockets, all five models produced during model building in MODELLER 9.12<sup>177</sup> were revisited. All of the models were now prepared, minimised and binding pockets were identified from receptor cavities using Discovery Studio 4.0.<sup>186</sup> The five models were then aligned, Figure 3.12 (a). Given the reasonably low sequence identity of the SMC5 head domain and 1XEX template<sup>224</sup>, it is plausible that the models may be inaccurate at the atomic level. However, the same adenosine binding pocket and neighbouring channel were observed in all the models, helping to improve confidence that they are real.

In addition, another model of the SMC5 head was produced using Phyre 2.0<sup>221</sup>, a protein fold recognition (threading) server. A crystal structure of the *Deinococcus radiodurans* DNA repair protein RecN (PDB ID: 4AD8)<sup>229</sup> was selected by the server as the template. The model was prepared and minimised and binding pockets identified from receptor cavities using both Discovery Studio 4.0<sup>186</sup> and Fpocket 2.0.<sup>165</sup> Fpocket identified 20 pockets in the new SMC5 model. Those at and nearest to the SMC N-terminal ATP binding site, which was identified by alignment of the model with the 1XEX template used previously<sup>224</sup>, are illustrated in Figure 3.12 (b). Again, the adenosine binding pocket and neighbouring channel were recognised, further improving confidence in their existence.



**Figure 3.12. Attempts to Improve Confidence in Structural Pockets.** (a) Alignment of five models of human SMC5 head. ATP from 1XEX template<sup>224</sup> in black. (b) Phyre 2.0<sup>221</sup> model of human SMC5 head with structural pockets identified by Fpocket 2.0.<sup>165</sup> SMC N- and C-terminal chains shown in yellow and blue with pockets highlighted and ATP from 1XEX template<sup>224</sup> in pink.

#### *Determining Whether the Pockets are Conserved in Other Proteins*

Those amino acids which comprise the ATP binding pocket and neighbouring channel were identified, considering both the 1XEX template<sup>224</sup> and SMC5/6 model. In Figure 3.13 these amino acids are compared with the corresponding residues in other SMC proteins, determined by multiple sequence alignment, Figure Apx.4.<sup>181,182</sup> The LIGPLOT<sup>374</sup> derived schematic diagram of protein-ligand interactions in 1XEX<sup>224</sup> was retrieved from canSAR v2.0<sup>160</sup>, Figure 3.14 (a). Those residues involved in binding the phosphate or adenosine moieties are specified in Figure 3.13. Scorecons ([http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/valdar/scorecons\\_server.pl](http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/valdar/scorecons_server.pl)) was used to calculate conservation scores for each position and these were coloured such that dark red indicates the poorest conservation and dark green the best.<sup>220</sup> Figure 3.14 (b) shows

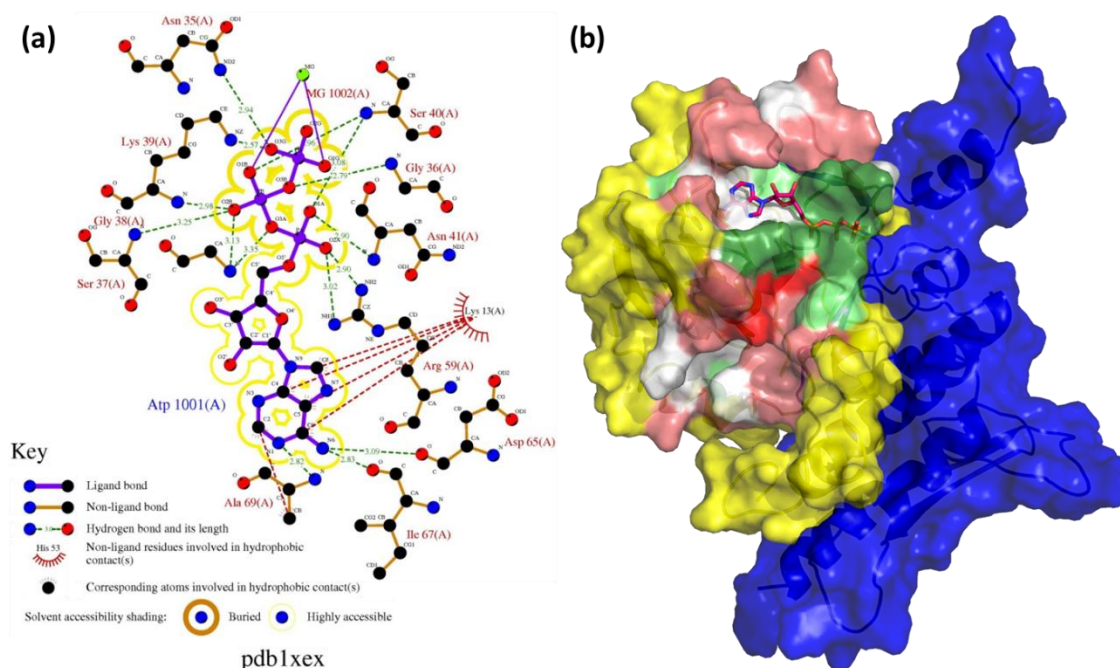


the model with the associated residues highlighted and coloured respectively. Residue conservation appears generally poor, with the exception of the phosphate binding residues, suggesting that a drug designed to target this region has the potential for reasonable specificity.

		Phosphate Binding								Adenosine Binding					Other Potential Binding Residues																		
Position on 1XEX		A35	A36	A37	A38	A39	A40	A41	A49	A13	A65	A67	A69	A14	A34	A42	A43	A44	A47	A50	A51	A52	A53	A54	A56	A57	A58	A60	A66	A68	A70	A71	
	SMC5	-	N	G	T	G	K	S	R	L	F	V	R	T	A	I	V	C	C	L	A	G	K	P	F	M	G	A	F	K	G	C	
	SMC6	-	N	G	S	G	K	S	R	-	G	V	D	N	N	V	L	T	I	L	G	G	R	A	A	T	N	G	F	K	G	Q	
	SMC2	-	N	G	S	G	K	S	N	R	K	D	V	K	S	L	I	L	D	C	L	G	I	S	N	S	Q	V	A	L	Y	N	G
	SMC3	-	N	G	S	G	K	S	N	R	R	R	A	L	S	R	F	F	Y	Q	L	S	D	E	F	H	-	L	P	L	L	H	E
	SMC4	-	N	G	S	G	K	S	N	R	K	V	I	N	S	P	V	I	D	L	F	G	Y	R	A	K	-	I	S	L	H	-	S
	SMC1b	-	N	G	S	G	K	S	N	R	K	E	I	G	S	P	V	M	D	S	M	G	E	K	I	N	-	L	V	L	H	A	H
	SMC1a	-	N	G	S	G	K	S	N	R	K	D	I	G	S	P	L	M	D	S	L	G	E	K	T	N	-	L	V	L	H	A	P
	RAD50	-	N	G	A	G	K	T	T	K	R	T	V	D	S	P	I	I	E	K	C	T	G	D	F	P	G	T	G	F	H	P	K
Conservation		1.00	1.00	0.71	1.00	1.00	0.84	0.59	0.89	0.48	0.27	0.68	0.34	0.71	0.44	0.60	0.53	0.41	0.29	0.58	0.59	0.38	0.48	0.33	0.31	0.06	0.36	0.41	0.71	0.50	0.29	0.30	

		Other Potential Binding Residues																															
Position on 1XEX		A126	A127	A130	A135	A136	A137	A -	A139	A140	A141	A142	A143	A144	A -	A145	A146	A147	A150	A151	B77	B80	B81	B88	B90	B92	B139	B140	B141	B142	B143	B144	
	SMC5	-	V	E	V	I	Q	V	G	L	C	Q	-	F	L	P	Q	D	K	E	F	L	L	Q	F	V	D	L	F	V	Y	N	G
	SMC6	-	L	I	L	I	Q	V	D	P	V	S	V	L	T	-	Q	E	M	Q	F	I	L	W	F	C	D	L	R	M	S	D	P
	SMC2	-	V	Q	F	L	N	V	N	P	H	F	L	I	M	-	Q	G	R	K	V	I	M	L	I	I	D	K	F	V	D	G	V
	SMC3	-	V	M	L	F	S	R	S	P	Y	Y	I	V	K	-	Q	G	K	Q	M	I	I	Q	F	L	D	K	F	R	N	K	V
	SMC4	-	V	G	L	I	D	L	D	N	R	F	L	I	L	-	Q	G	E	Q	I	V	L	H	L	F	D	Y	K	T	Y	N	I
	SMC1b	-	Y	I	L	I	I	V	K	Q	N	C	L	V	F	-	Q	G	T	S	I	L	V	H	F	V	D	Y	P	E	Y	D	D
	SMC1a	-	Y	S	L	I	L	I	K	R	N	F	L	V	F	-	Q	G	A	S	I	L	I	H	F	V	D	Y	P	E	Q	G	D
	RAD50	-	D	R	I	V	S	K	A	L	N	N	V	I	F	C	Q	E	D	-	W	R	E	T	I	A	D	K	K	N	I	D	Q
Conservation		0.41	0.29	0.64	0.65	0.33	0.46	0.39	0.39	0.36	0.36	-	0.59	0.39	0.01	1.00	0.64	0.34	0.35	0.42	0.50	0.47	0.37	0.62	0.42	1.00	0.38	0.32	0.32	0.33	0.50	0.30	

**Figure 3.13. Conservation of SMCs.** Scoring conservation of potential binding site residues in human SMCs and Rad50.

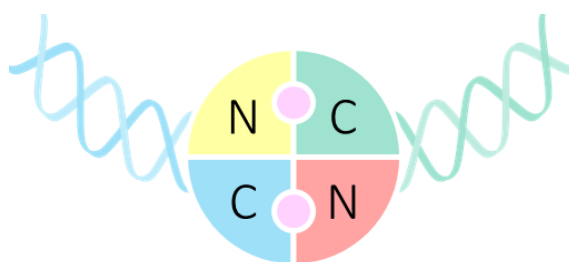


**Figure 3.14. Attempts to Determine Pocket Conservation in Other Human SMC Proteins.** (a) Schematic Diagram of Protein-Ligand Interactions of 1XEX<sup>224</sup> generated by LIGPLOT.<sup>374</sup> (b) Model of human SMC5 head with residues at potential binding pockets coloured by conservation. SMC N- and C-terminal chains shown in yellow and blue and ATP from 1XEX template<sup>224</sup> in pink.

Analysis of whether similar pockets exist in other protein families was not undertaken, due to the manner and extent of work required to achieve this.

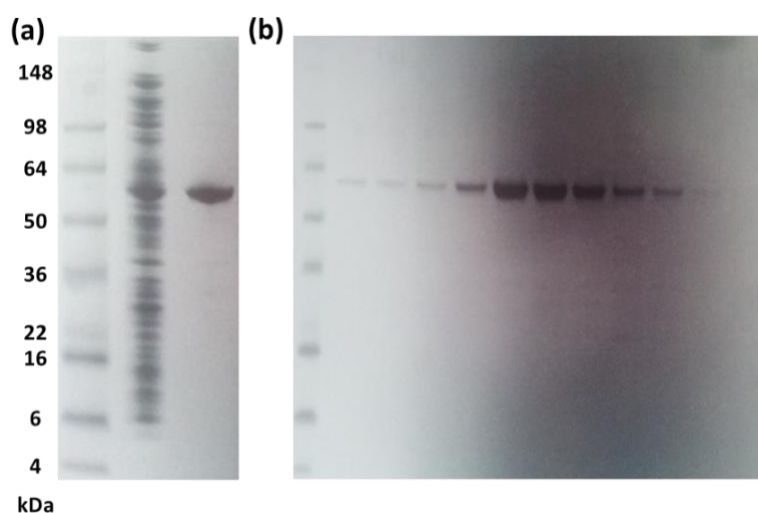
#### *Experimental Assessment of SMC5 Head Domain*

As discussed, ATP binding to SMC5 and SMC6 occurs at the interface between the head domains of the two proteins to form a nucleotide-sandwiched dimer, in which the two ATP molecules each interact with both subunits, as illustrated in Figure 3.15. Consequently, the SMC5 or SMC6 head domain alone comprises only one half of each of the ATP binding sites. Given that purified SMC6 protein is currently unavailable it was necessary to assess the ability of the head domain of SMC5 to bind ATP moieties in its absence.



**Figure 3.15. Illustration of SMC5/6 Head Domains with ATP Bound.** ATP shown in pink, SMC5 N- and C-terminal chains shown in yellow and blue respectively and SMC6 in red and green.

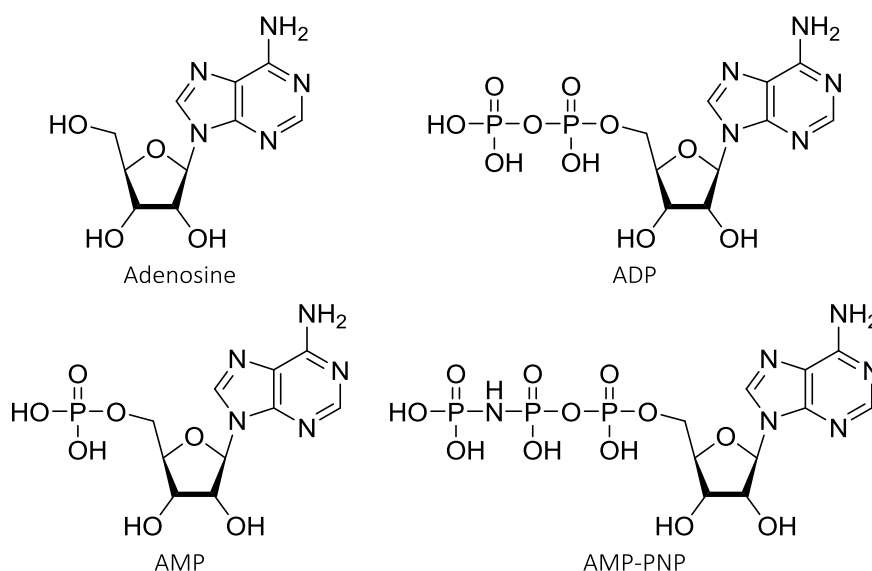
SMC5 head domain protein, generously expressed by Dr Aaron Alt (Genome Damage and Stability Centre, University of Sussex), was purified using immobilised metal affinity chromatography (IMAC) and size exclusion chromatography (SEC), from frozen cells, Figure 3.16. The construct, AA5, comprised of the N- and C-terminal regions of SMC5 joined by a short linker.



**Figure 3.16. SMC5 Head Domain Protein Purification.** (a) SDS-PAGE gel showing composition of wash and elute fractions from immobilised metal affinity chromatography. (b) SDS-PAGE gel of fractions obtained by size exclusion chromatography. Concentrated protein band corresponds to SMC5 head domain protein.

Isothermal titration calorimetry (ITC) was employed in an attempt to measure the corresponding heats of binding of adenosine, AMP, ADP and AMP-PNP, a non-hydrolysable ATP analogue,

Figure 3.17. The results of experiments conducted at 1.5 mM compound and 60  $\mu$ M protein are shown in Figure 3.18. On observation of binding you might expect to detect energy changes of over 10 kcalmol<sup>-1</sup> of injectant. In these experiments however, the measured 'heat of binding' was comparable with the heat of dilution, indicating that none of the ATP analogues bound with the SMC5 head domain protein. No binding curves could be fit to the resulting data.



**Figure 3.17. Chemical Structures of ATP Moieties.** Adenosine, AMP, ADP and AMP-PNP.

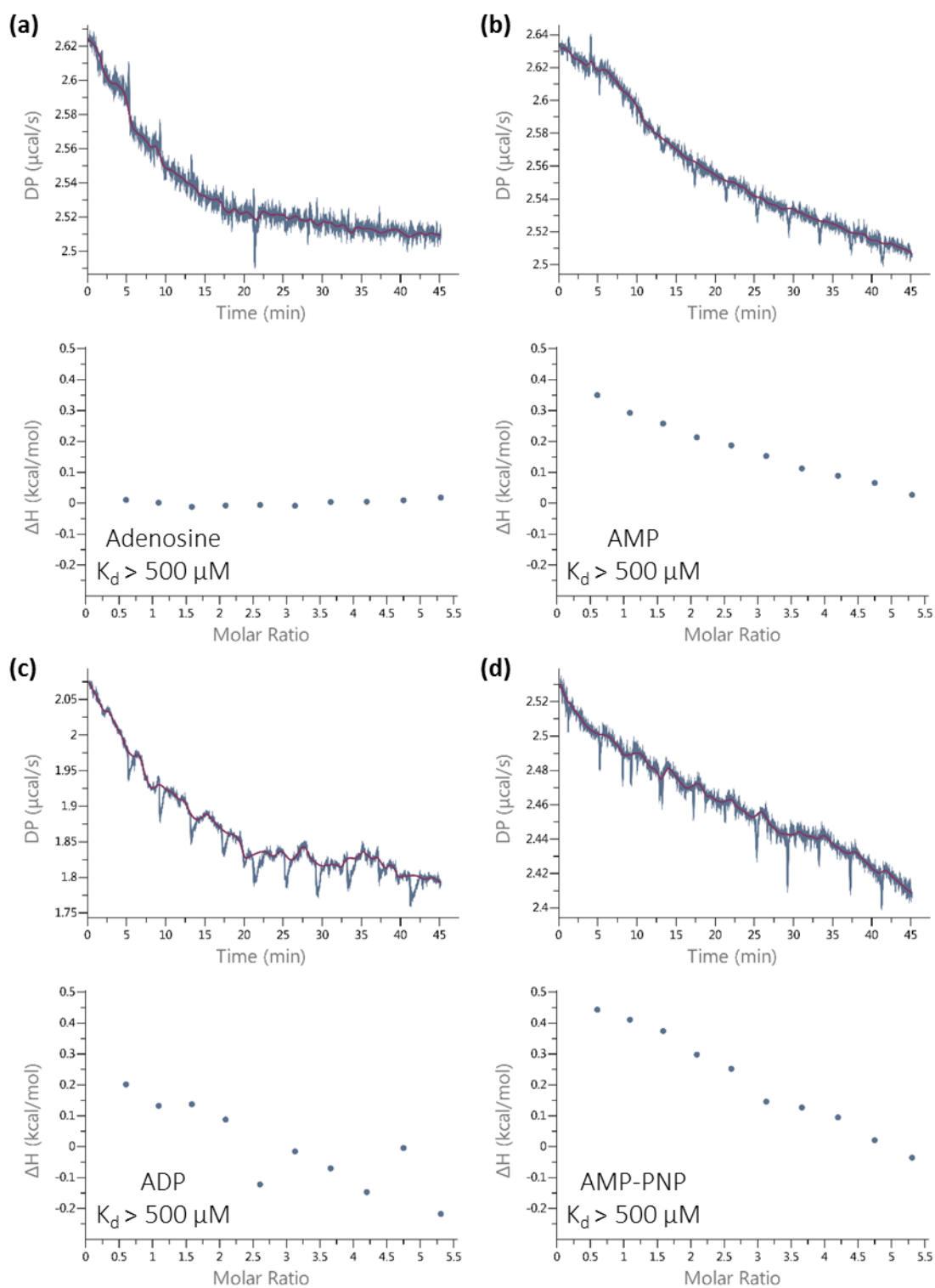
From these results it appears that the head domain of SMC5 is unable to bind ATP moieties in the absence of that of SMC6. As described previously, one feature of a favourable drug target is that a robust assay can be developed against the protein, which is important in enabling screening and drug development. It was therefore determined that while purified SMC6 protein remained unavailable, the head domain of the SMC5/6 complex, like the hinge, did not constitute a good drug target.

### 3.2.2 Non-SMC Elements

In addition to the two core SMC proteins, the human SMC5/6 complex contains four non-SMC elements, NSE1, 2, 3 and 4.<sup>342</sup> In order to obtain a complete picture of the potential druggability of the complex, it was important to also assess these.

DELTA-BLAST<sup>180,183</sup> searches identified suitable structures on which to model NSE1, NSE2 and NSE3 but no structures were available which showed sufficient sequence similarity to NSE4 to enable modelling. In addition an attempt to use PHYRE 2.0<sup>221</sup> to build a model of NSE4 was also unsuccessful due to the lack of suitable structures available in the PDB. MODELLER 9.12<sup>177</sup> was used to model NSE1, NSE2 and NSE3 on the most suitable available crystal structures. NSE1 was modelled on the *Homo sapiens* NSE1-MAGEG1 complex (PDB ID: 3NW0, chain A)<sup>225</sup>, NSE2 was modelled on the *S. cerevisiae* MMS21 (PDB ID: 3HTK, chain C)<sup>226</sup>, and NSE3 was modelled on the

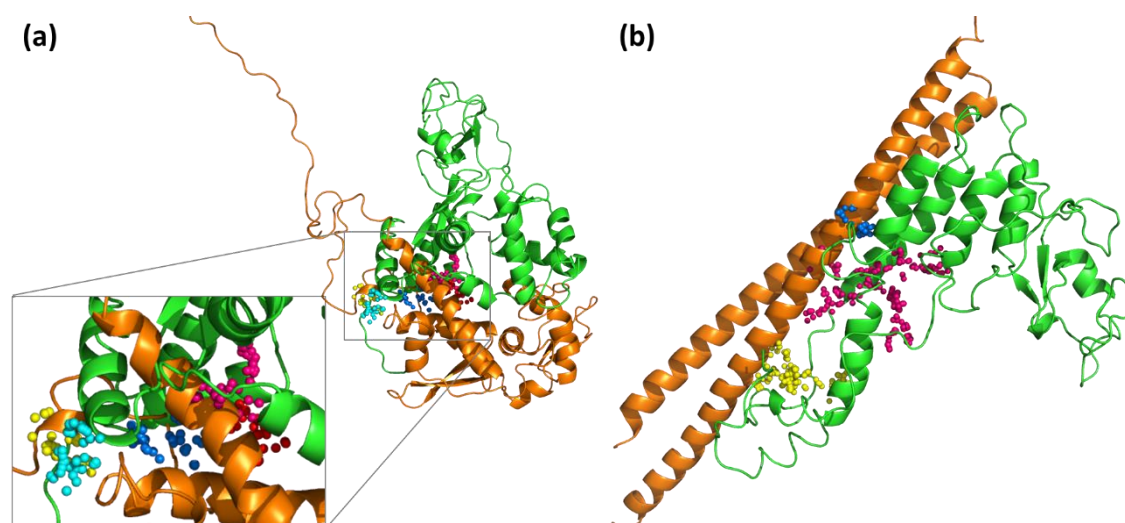
*Homo sapiens* MAGEA4 (PDB ID: 2WA0)<sup>227</sup>. The models were prepared and minimised, the binding pockets identified from receptor cavities using both Discovery Studio 4.0<sup>186</sup> and Fpocket 2.0<sup>165</sup> and the associated Fpocket druggability scores<sup>170</sup> calculated.



**Figure 3.18. ITC Binding Data for Interaction Between SMC5 Head Domain Protein and ATP Analogues.** Isotherms of raw titration heat shown above and normalised binding heat shown below. No binding was observed for any of the ATP analogues. (a) 1.5 mM adenosine titrated into 60  $\mu\text{M}$  protein. (b) 1.5 mM AMP titrated into 60  $\mu\text{M}$  protein. (c) 1.5 mM ADP titrated into 60  $\mu\text{M}$  protein. (d) 1.5 mM AMP-PNP titrated into 60  $\mu\text{M}$  protein.



In the SMC5/6 complex NSE1 and NSE3 are known to interact with each other.<sup>342</sup> PDB entry 3NW0<sup>225</sup> depicts the structure of the two proteins in a bound conformation but unfortunately is of poor quality, especially with respect to the NSE3 subunit, so could not be used to model NSE3. Instead, while NSE1 was modelled on the 3NW0<sup>225</sup>, to reasonably good effect, NSE3 was modelled on 2WA0.<sup>227</sup> The resulting NSE3 model was then aligned to 3NW0, chain B<sup>225</sup>, in an attempt to observe the PPI interface. 10 pockets were identified in each of the NSE1 and NSE3 models using Fpocket 2.0.<sup>165</sup> Those at and nearest to the site of interaction, are illustrated in Figure 3.19 (a). On both proteins, cavities were identified at the PPI interface, although the associated druggability scores, displayed in Table 3.4, were low. The poor quality of the available X-ray crystal structures used in modelling results in a high likelihood of inaccuracy in these models, making them unsuitable for use in further analysis. This investigation was therefore abandoned, but should be revisited if improved structures become available.



**Figure 3.19. Modes of Human NSE1, NSE3 and NSE2, with Structural Pockets from Fpocket 2.0.**<sup>165</sup> (a) Models of NSE1 and NSE3 aligned to display site of interaction. NSE1 shown in green and NSE3 in orange with pockets highlighted. (b) Model of NSE2 bound to SMC5. NSE2 shown in green and SMC5 arm in orange with pockets highlighted.

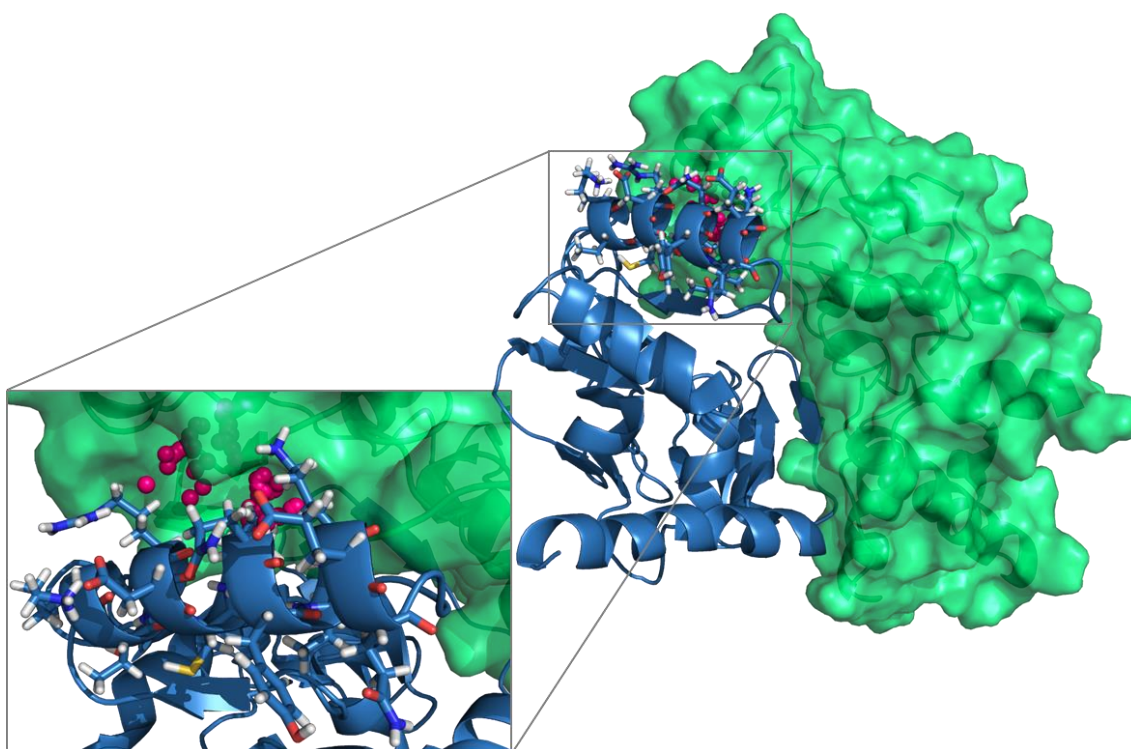
NSE2 binds to the SMC5/6 dimer independently through interaction at the SMC5 coiled-coil arm.<sup>346</sup> This interaction is depicted in PDB entry 3HTK.<sup>226</sup> 14 pockets were identified in the NSE2 model by Fpocket 2.0.<sup>165</sup> Those at and nearest to the interface with the SMC5 arm are illustrated in Figure 3.19 (b) and the corresponding druggability scores are displayed in Table 3.4. One pocket, highlighted in yellow, has an associated druggability score of 0.551, which is higher than any seen in the SMC5/6 head domain or other NSEs and might suggest that the cavity would make a viable target site. However, given that in humans NSE2 is not thought to be necessary for the stability of the SMC5/6 complex<sup>343</sup>, before further time and resources is committed to a drug design effort at this position, it would be necessary to determine whether targeting NSE2 affords the desired cellular effect. Point mutation experiments could be utilised toward this aim. Accordingly, this analysis should also be revisited if more data becomes available.

**Table 3.4. Druggability Prediction Scores from Fpocket 2.0.<sup>170</sup>**

	Pocket	Colour	Druggability Score
Human NSE1 Model	4	Pink	0.076
	8	Yellow	0.014
	10	Blue	0.385
Human NSE3 Model	4	Red	0.016
	8	Cyan	0.009
Human NSE2 Model	1	Pink	0.468
	3	Yellow	0.551
	10	Blue	0.023

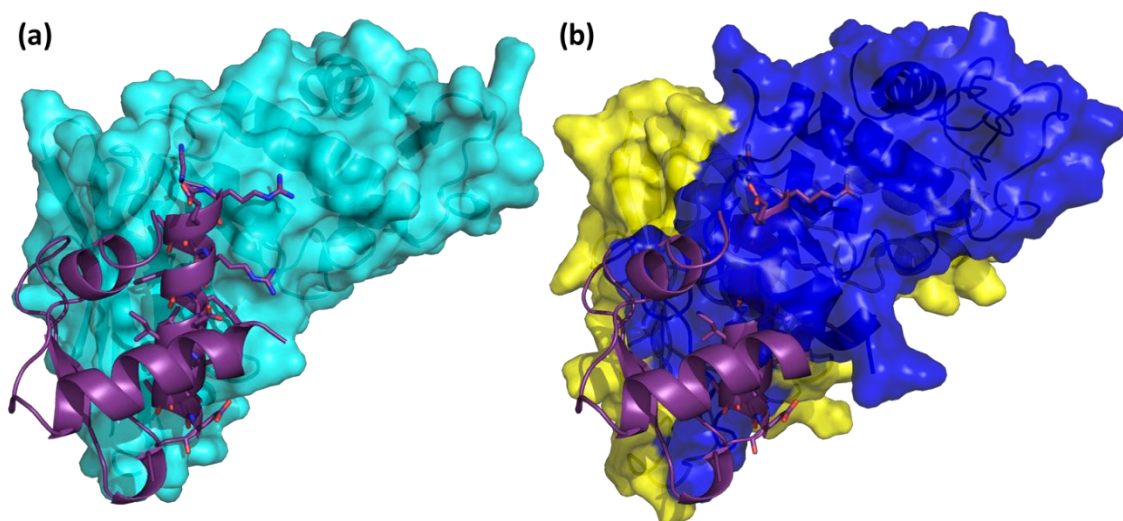
### 3.2.3 Cohesin

In addition to their primary roles in sister chromatid cohesion and chromosome condensation respectively, cohesin and condensin both have established roles in DNA damage response.<sup>251</sup> Consequently, like SMC5/6, these protein complexes represent biologically desirable drug targets. Relevant data on cohesin and condensin was collected from a variety of online sources and databases and information was gathered on the availability of resources and potential for collaboration with other groups. Cohesin was prioritised, owing to a greater amount of available information, potential to obtain X-ray crystal structures (Ramón-Maiques, S., Spanish National Cancer Research Centre, personal communication) and potential for development of an in-cell assay.<sup>272</sup>



**Figure 3.20. Model of Human Cohesin Hinge Domain.** SMC1 shown in blue and SMC3 in green with compelling structural pocket identified by Fpocket 2.0.<sup>165</sup> highlighted in pink.

The most suitable available crystal structures were identified using a DELTA-BLAST<sup>180,183</sup> search against the PDB, and the SMC1/3 hinge and head domains were modelled respectively on the *Mus musculus* SMC1/3 hinge domain (PDB ID: 2WD5)<sup>228</sup>, and the *Pyrococcus furiosus* SMC ATPase (PDB ID: 1XEX)<sup>224</sup> using MODELLER 9.12.<sup>177</sup> The models were prepared and minimised, the binding pockets identified from receptor cavities using Discovery Studio 4.0<sup>186</sup> and Fpocket 2.0<sup>165</sup>, and the corresponding Fpocket druggability scores were calculated.<sup>170</sup> In general, the models were similar to those of the SMC5/6 hinge and head domains. Interestingly though, one pocket identified in the SMC3 hinge domain at the SMC1-SMC3 PPI interface corresponds to the binding position of a helix in SMC1, as illustrated in Figure 3.20. The Fpocket druggability score for this cavity is 0.192 suggesting that the site is not druggable. However, as discussed earlier, the druggability prediction scores may not be truly reflective of the druggability of solvent exposed cavities such as this at the PPI interface. Use of helix mimetics to target PPI interfaces is well documented<sup>375</sup> and this may provide a potential starting platform for drug design.



**Figure 3.21. Analysis of the SMC1-SCC1 Interaction Interface.** (a) *S. cerevisiae* SMC1-SCC1 interaction interface (PDB ID: 1W1W).<sup>233</sup> SMC1 shown in cyan and SCC1 in purple. (b) Model of human SMC1 head domain aligned to *S. cerevisiae* SMC1. SMC N- and C-terminal chains shown in yellow and blue with SCC1 in purple. Alignment results in obscuring of the helix binding groove.

As well as 3D structures of the hinge and head domains, there is a published crystal structure of the *S. cerevisiae* SMC1-SCC1 interaction interface (PDB ID: 1W1W).<sup>233</sup> Such information was not available for SMC5/6 and was interesting as any proteins or small molecules bound to the SMC proteins might provide inspiration for potentially druggable sites, along with corresponding pharmacophores. Binding pockets were identified from receptor cavities using Fpocket 2.0<sup>165</sup> and the associated Fpocket druggability scores were calculated.<sup>170</sup> 11 pockets were identified in SMC1 (PDB ID: 1W1W, chain A).<sup>233</sup> Unfortunately none of these were situated at the SMC1-SCC1 interface, likely due to the extensive, flat and lipophilic nature of the interaction. Furthermore, alignment of the *S. cerevisiae* SMC1 with the model of the human SMC1/3 heads was

unsuccessful. The majority of the helix binding groove became completely obscured, as shown in Figure 3.21, probably as a combined result of inaccuracies in modelling and alignment.

While cohesin looks appealing as a potential future target for development of tool inhibitors, any progress will be highly dependent on the level of interest of potential collaborators and availability of resources.

### 3.3 Discussion

The roles of the SMC complexes in regulating chromosome architecture and organisation are well known, but not entirely well understood. SMC5/6 is particularly mysterious, with both the essential function and mechanism of action unexplained, despite considerable investigative effort. In addition, SMC5/6 has been shown to play a prominent role in DNA repair<sup>254–256</sup> and in the ALT pathway<sup>361</sup>, as well as representing a potential opportunity to design novel cancer therapeutics via a synthetic lethality approach. SMC5/6 was therefore identified as a highly desirable target for tool inhibitor development. Such chemical probes are anticipated to enable enhanced biological investigation and could initiate drug discovery projects. Before commencing any chemistry, it is prudent to use druggability assessment methods to determine the viability of the protein as a target. This enables promising targets to be prioritised quickly and cheaply and goes a small way toward helping improve R&D productivity.<sup>14</sup>

In this work, the human SMC5/6 hinge was modelled successfully on the *S. pombe* SMC5/6 hinge domain (PDB ID: 5MG8).<sup>223</sup> *S. pombe* SMC5 residues Ser-610 and Tyr-612, which are necessary to stabilise the interaction at the hinge<sup>223</sup>, were found to be conserved in humans (Ser-624 and Tyr-626). Furthermore, these residues were similarly positioned at the interface with SMC6, suggesting a key role in preserving the stability of the hinge interaction. Unfortunately, at this time the instability of SMC6 prohibited development of an assay against this protein, making it impossible to test drugs designed to mimic these SMC5 residues. Instead, it was necessary to identify potential binding pockets from receptor cavities on SMC5. However, it was discovered that the interface was composed only of  $\beta$ -sheets, with no discernible structural pockets, affording little potential for a rationally designed PPI inhibitor. It therefore seemed that the SMC5 hinge domain was undruggable. It is advised that if a human crystal structure is determined, if proteins or small molecules are identified that bind at the SMC5 hinge, or if stable SMC6 can be produced, the SMC5/6 hinge should be reinvestigated as a target.

With opportunities at the hinge exhausted, focus shifted to the SMC5/6 head domain. The human SMC5/6 head was successfully modelled on the *Pyrococcus furiosus* SMC ATPase (PDB ID: 1XEX).<sup>224</sup> SMC5 was the preferred target. As experienced at the hinge, the SMC6 head domain

was more unstable than the SMC5, and attempts to prepare and purify it had been unsuccessful. Potential binding pockets were identified at and near to the SMC5 N-terminal ATP binding site. The associated druggability scores<sup>170</sup> suggested that these pockets were undruggable. However, it was judged that this could be caused by the solvent exposed nature of the pockets, or the polar environment required to accommodate the ATP phosphate residues, and therefore was potentially not truly reflective of the druggability. In attempt to validate the two pockets identified around the adenosine binding site, a second model of the SMC5 head was produced, this time using an energetics approach rather than alignment. Equivalent binding sites were observed, improving confidence that the pockets exist. A sequence based approach was used to compare residues that could potentially comprise a druggable pocket around the SMC5 N-terminal ATP binding site with the corresponding residues in other human SMC proteins and RAD50. Conservation was generally poor, with the exception of the phosphate binding residues. This suggested that a drug targeted at this site would have the potential to offer reasonable selectivity for SMC5 over other SMC family members. Having identified the potentially druggable sites, ITC was used to assess the ability of the SMC5 head to bind ATP derivatives in the absence of SMC6. Unfortunately, no binding could be measured between the SMC5 head domain and adenosine, AMP, ADP or AMP-PNP. As such, while purified SMC6 head protein remains unavailable, the SMC5/6 head domain also appears undruggable.

Next, investigations advanced to the non-SMC elements. NSE1, 2 and 3 were modelled on the *Homo sapiens* NSE1-MAGEG1 complex (PDB ID: 3NW0, chain A)<sup>225</sup>, *S. cerevisiae* MMS21 (PDB ID: 3HTK, chain C)<sup>226</sup>, and *Homo sapiens* MAGEA4 (PDB ID: 2WA0)<sup>227</sup> respectively. Although potential cavities were identified at the NSE1-NSE3 PPI interface, the poor quality of the available X-ray crystal structures resulted in a high likelihood of inaccuracy in the models, making them unsuitable for further analysis. This target was abandoned but should be revisited if improved crystal structures become available. Potentially druggable pockets were identified at the NSE2-SMC5 PPI interface. However, given that in humans NSE2 is not thought to be necessary for the stability of the SMC5/6 complex<sup>343</sup>, it was concluded that before time and resources were committed to any drug design effort at this position, point mutation experiments should be utilised to determine whether targeting NSE2 generates the desired response.

Having failed to identify a satisfactory target protein within the SMC5/6 complex, attention was redirected to cohesin. Cohesin was selected over condensin due to the abundance of available information, and potential to develop an in-cell assay<sup>272</sup> and obtain X-ray crystal structures via collaboration. The human cohesin hinge and head domains were successfully modelled on the *Mus musculus* SMC1/3 hinge domain (PDB ID: 2WD5)<sup>228</sup>, and the *Pyrococcus furiosus* SMC

ATPase (PDB ID: 1XEX)<sup>224</sup> respectively. A pocket was identified in the SMC3 subunit at the PPI interface between the hinge domains, which corresponded to the binding position of a helix in the SMC1. Although the Fpocket druggability score<sup>170</sup> for this cavity indicated that it was not druggable, it seemed that this might not truly reflect the druggability of such a solvent exposed pocket. The site could offer the potential for drug design utilising helix mimetics. Analysis of PDB entry 1W1W<sup>233</sup>, which depicts the *S. cerevisiae* SMC1-SCC1 interaction interface, identified a lack of binding sites at this interface, likely due to the extensive, flat and lipophilic nature of the interaction. Cohesin could represent a promising future target for development of tool inhibitors. However, for this project it was concluded that the necessary collaborations and resources could not be assembled.

In summary, at the time this work was completed, the SMC5/6 and cohesin complexes were established as not druggable, within the restraints of accessible resources. In the future, if further data and biological resources become available, many of the proteins would be worth reinvestigating. However, within the scope of this project it was not considered a wise investment to attempt design and characterisation of tool inhibitors.

## 4 Assessing the Druggability of Three Targets that are Synthetic Lethal with BAF180

BAF180, gene name PBRM1, is one of the defining subunits of PBAF, a SWI/SNF complex associated with chromatin formation.<sup>376</sup> Mutations in PBRM1 have frequently been identified in a variety of human cancers.<sup>377</sup> Most recently, PBRM1 was identified as a major clear cell renal cell carcinoma (ccRCC) cancer gene, exhibiting truncating mutations in 41% of samples of the protein coding exome in a series of primary ccRCCs.<sup>1</sup> It is therefore the second most frequently mutated gene in ccRCC.<sup>376</sup>

Unique features of tumours, such as this high frequency of BAF180 mutation, can be exploited by targeted therapies using a synthetic lethality approach and therefore highlight suitable targets for chemotherapeutic drug development. Hopkins *et al.* utilised a novel screening technique to identify several genes that are synthetic lethal with BAF180.<sup>3</sup> In this chapter, the druggability of proteins expressed by three of these genes, RNF4, ASF1A and GCN5/KAT2A, was assessed using computational techniques. GCN5/KAT2A was determined to be the most druggable of the three proteins, resulting in initiation of tool inhibitor development, as described in Chapters 5 and 6.

### 4.1 Introduction

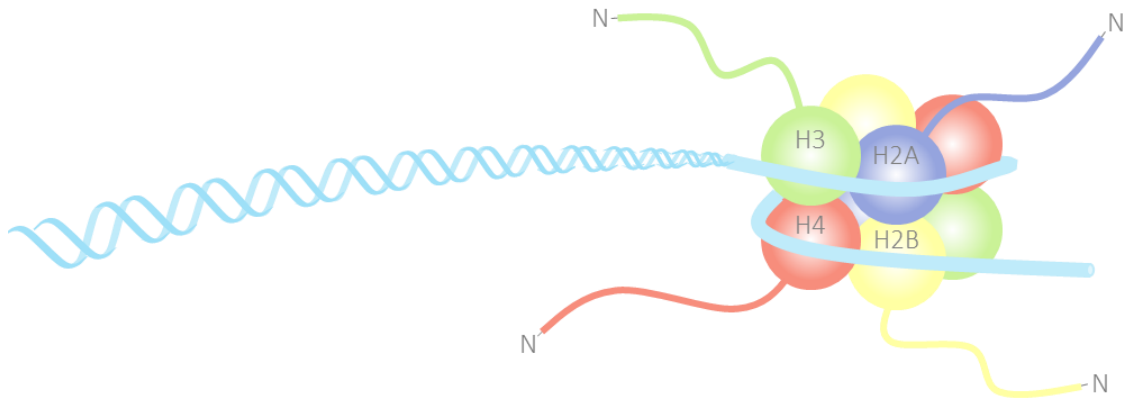
#### 4.1.1 Chromatin Remodelling Complexes

As revealed in Chapter 3.1, higher-level chromosome condensation is not completely understood. Opinion is split between hierarchical folding or radial loop models, as well as whether protein scaffolds exist at chromosome axes to constrain the structure.<sup>378</sup> In contrast, the primary level of DNA organisation is well studied and understood, to the extent that the structure has been resolved at 1.9 Å resolution (PDB ID: 1KX5).<sup>379</sup>

The charged DNA polymer is compacted into nucleosomes, through association with an octameric complex of highly basic histone proteins (H2A, H2B, H3 and H4), as shown in Figure 4.1. These nucleosomes are the fundamental repeating unit of chromatin and generally occur every 157-240 base pairs.<sup>379</sup> The chromatin is then successively folded by some means to afford higher-order and remarkably well condensed structures.<sup>380</sup> Although less is known about the nature of these more folded structures, it is clear that the degree of compaction directly influences the activity of the associated DNA in transcription, replication and recombination.<sup>381</sup>

ATP-dependent chromatin remodellers are key components of the cellular machinery required for functional DNA packing. They use energy from ATP hydrolysis to remodel nucleosomes and

thereby modulate transcription.<sup>382</sup> Notably, such genes are among those most frequently disrupted in human tumours. Studies show that 19-20% of human malignancies exhibit defects in the subunits of two chromatin remodelling complexes, BAF (BRG1- or BRM-associated factor) and PBAF (polybromo-associated BAF).<sup>383,384</sup> These are the focus of this discussion.



**Figure 4.1. DNA Packaging into Nucleosomes.** DNA is wrapped around an octameric core of histone proteins.

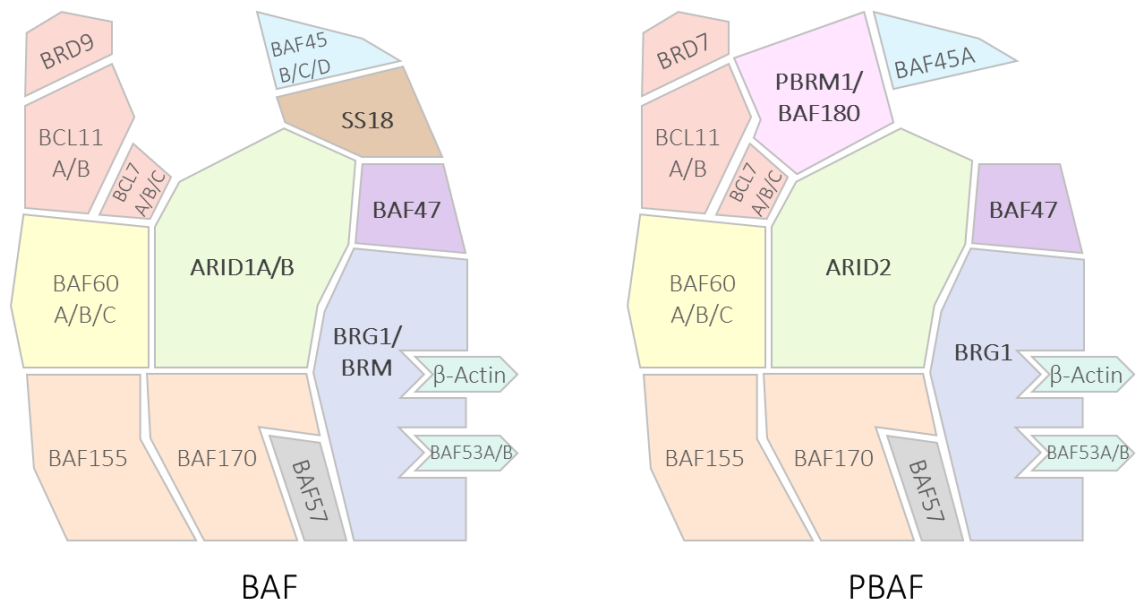
The yeast SWI/SNF (SWItch/Sucrose Non-Fermentable) chromatin remodeller was discovered independently in screens for mutations which hinder mating type switching<sup>385</sup>, or sucrose fermentation activation.<sup>386</sup> The genes identified were later shown to be part of the same complex.<sup>387,388</sup> SWI/SNF complexes function by modifying the position, phasing, stability or histone content of nucleosomes.<sup>389</sup> Structural remodelling enables the complexes to mobilise nucleosomes either by sliding or by catalysing the ejection and insertion of histone octamers, to generate sites more accessible to DNA binding factors.<sup>390</sup> Cells with mutant SWI/SNF subunits exhibit disrupted chromatin structure and impaired gene expression, causing an array of phenotypic defects, as well as increased sensitivity to DNA damaging agents.<sup>376</sup> This sensitivity may derive from the roles of SWI/SNF complexes in nucleotide excision<sup>391</sup> and homologous recombination<sup>392</sup> repair pathways.

The SWI/SNF complexes comprise a highly related family of multisubunit complexes.<sup>382</sup> The mammalian analogues are the aforementioned BAF, also known as SWI/SNF-A; and PBAF, also known as SWI/SNF-B. These complexes are composed of a central ATPase, either BRG1 or BRM, various evolutionarily conserved BAF subunits, and several subunits specific to mammalian complexes, including BCL7A/B/C and BCL11A/B. The two complexes differ in that BAF contains ARID1A/B, BRD9, BAF45B/C/D and SS18, where PBAF contains ARID2, BRD7, BAF45A and PBRM1/BAF180, as presented in Figure 4.2.<sup>376</sup> These subunits are thought to contribute to specific targeting, assembly and regulation of the complexes.<sup>382</sup>

The BAF complex is heavily implicated in mammalian development. It regulates a variety of biological processes including self-renewal and pluripotency in embryonic stem cells, cardiac



development and neural differentiation<sup>393</sup>, and is observed to localise at developmental enhancers.<sup>394</sup> Additionally, its activity is coupled to transcription factor binding.<sup>395</sup> PBAF was first identified in a screen for proteins that play an essential role in ligand-mediated transcription.<sup>396</sup> This complex accumulates at the kinetochores of mitotic chromosomes, suggesting a role in cell division<sup>397</sup>, and performs in cell differentiation.<sup>398</sup> In addition, analogously to SWI/SNF, there is evidence that BAF and PBAF function in DNA repair pathways, including nonhomologous end-joining and homologous recombination repair.<sup>399</sup> Correspondingly, it has long been speculated that BAF subunits are tumour suppressors, as these core subunits are recurrently absent or mutated in immortalised cell lines.<sup>400,401</sup> Indeed, BAF and PBAF have now been confirmed as legitimate tumour suppressors, and specific inactivating mutations have been frequently identified in a number of the subunits in various human cancers.<sup>376</sup>



**Figure 4.2. SWI/SNF Complexes BAF and PBAF.** PBRM1 unique to PBAF and SS18 to BAF. Subunits with important roles in malignancy labelled in bold. Figure adapted from Hodges *et al.*, 2016 and Kadoch & Crabtree, 2015.<sup>376,402</sup>

#### 4.1.2 BAF180

BAF180, gene name PBRM1 (Protein polybromo-1), is one of the defining subunits of PBAF, and is the focus of the synthetic lethality drug development approach investigated in this chapter.<sup>376</sup> It contains six tandem bromodomains, two bromo-adjacent homology (BAH) domains, and a high-mobility group (HMG). Bromodomains are roughly 100 amino acids in length and typically serve to recognise and selectively bind acetylated lysines from histones, enabling targeting of complexes to precise chromatin sites. However, the specific role of these bromodomains in targeting PBAF remains unclear.<sup>376</sup> The BAH domains are regions of approximately 130 amino acids, which act as protein-protein interaction modules, potentially acting to anchor the BAF180 subunit within the PBAF complex and recruit effector proteins. Finally, the HMG domain, or

HMG-box, is roughly 80 amino acids in length and binds into the minor groove of DNA, which is likely to alter histone-DNA interactions and thereby control genetic functions. Collectively it therefore appears that BAF180 serves as an important PBAF subunit, coordinating several roles vital to the function of chromatin remodelling complexes.<sup>403</sup>

Mutations in PBRM1 are associated with a variety of human cancers, such as breast cancers<sup>377</sup> and clear cell renal cell carcinoma, with a recent study identifying truncating mutations in 41% of primary ccRCC samples<sup>1</sup>, making it the second most frequently mutated gene in ccRCC.<sup>376</sup> In the case of ccRCC, inactivation of PBRM1 often coincides with mutation of tumour suppressor VHL (von Hippel-Lindau). This is likely principally due to their close proximity on chromosome arm 3p, such that focal or arm-level deletions affect both genes simultaneously. However, inactivating point mutations of PBRM1 also regularly coincide with VHL mutation, suggesting that joint inactivation of these genes may be beneficial to the cancer.<sup>404</sup>

Although PBAF primarily regulates transcription, it is unclear if this is what drives tumourigenesis in cells deficient in BAF180.<sup>405</sup> BAF180 has a number of specifically defined roles in maintaining genome stability and preventing tumourigenesis.<sup>399</sup> Firstly, it appears to function in repriming stalled replication forks at sites of DNA damage.<sup>406</sup> Replication past damaged DNA is regulated by PCNA ubiquitination. Following damage-inducing UV-irradiation, BAF180 deficient cells present reductions in both ubiquitinated PCNA and chromatin-associated unmodified PCNA, as well as reduced fork progression. It has therefore been suggested that PBAF, via BAF180, promotes repriming of replication downstream of replication forks blocked by damage, by recruiting and ubiquitinating PCNA. In the absence of BAF180, replication is therefore slowed.<sup>406</sup> In accordance with this hypothesis, it has recently been shown that the BAF180 BAH domains are necessary and sufficient to promote PCNA ubiquitination.<sup>407</sup> Secondly, it has been shown that BAF180 is necessary for centromeric sister chromatid cohesion, which could help to prevent tumourigenesis.<sup>405</sup> BAF180 depletion is not associated with misregulated expression of the SMC complex cohesin, which seemingly indicates a transcription-independent role for PBAF. This is consistent with the observation that PBAF accumulates at the kinetochores of mitotic chromosomes.<sup>397</sup> As well as bestowing accurate chromosome segregation, cohesion is important for DNA damage repair by HR, and mammalian cells with defects in cohesion are hypersensitive to DNA-damaging agents.<sup>408</sup> The loss of BAF180 similarly sensitises cells and therefore it is suggested that in the absence of BAF180, loss of sister chromatid cohesion results in impaired HR-dependent DNA repair.<sup>405</sup>

Unique mutations that are associated with tumourigenesis can potentially be exploited by targeted therapies, and are currently a key focus in cancer research efforts using synthetic

lethality approaches. In this instance, the presence of mutated PBRM1 in cancer cells but not in normal cells could create an opportunity to selectively kill the tumour cells by mimicking the effect of a synthetic lethal mutation with a drug. Indeed, this prominent mutation of the PBRM1 gene has been highlighted as a promising opportunity to target ccRCC.<sup>2</sup>

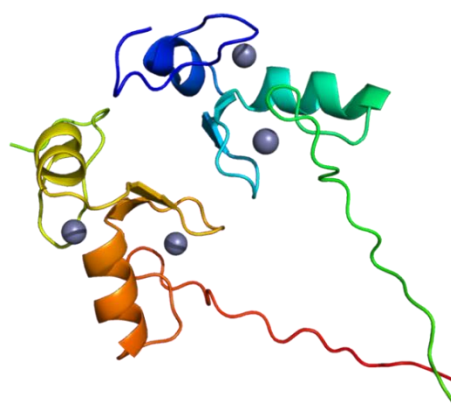
Hopkins *et al.* utilised a novel method of screening to identify genes synthetic lethal with BAF180.<sup>3</sup> The screening technique was based on shRNA and siRNA double knockdown; co-culturing isogenic cell lines expressing experimental or control shRNA with individual fluorescent tags and applying a library of siRNAs to diagnose synthetic lethal interactors, Figure 1.11.<sup>3</sup> Three of the proteins expressed by genes identified by Hopkins *et al.* were selected as biologically desirable targets for tool inhibitor development. RNF4, ASF1A and GCN5/KAT2A therefore became the focus of a computational druggability study to assess their suitability as targets of chemotherapeutic agents that exploit the synthetic lethality with BAF180.

#### 4.1.3 RNF4

RNF4 (RING finger protein 4), also known as SNURF (small nuclear RING finger protein), is a RING-domain-containing, SUMO-targeted E3 ubiquitin ligase<sup>409</sup>, which exists as a homodimer via its RING-type zinc finger domain, Figure 4.3.<sup>234</sup> RNF4 mediates SUMO-targeted polyubiquitination to designate various poly-SUMO-modified proteins for degradation at the proteasome.<sup>409</sup> This response is exploited in the treatment of acute promyelocytic leukaemia (APL) using arsenic.<sup>410</sup> After exposure of cells to arsenic, the oncogenic PML-RAR $\alpha$  (retinoic acid receptor alpha) fusion protein, commonly observed in APL, undergoes rapid SUMO modification. Subsequent RNF4-mediated ubiquitination results in the proteasomal degradation of PML-RAR $\alpha$ .<sup>410</sup> RNF4 is also responsible for ubiquitination and destabilisation of PEA3 (polyomavirus enhancer activator 3), promoting transactivation, possibly by inducing recycling<sup>411</sup>, and of CENPI (centromere protein I), to control vertebrate kinetochore assembly.<sup>412</sup> Similarly, it is implicated in regulating the cellular response to both hypoxia<sup>413</sup> and heat shock<sup>414</sup> via ubiquitination of SUMO-conjugated HIF2 $\alpha$  (hypoxia-inducible factor 2 alpha) and PARP1 respectively. Alternatively, RNF4 may have a more direct role in transcriptional regulation. For example, it has also been found to interact with TRPS1 to negatively regulate its transcriptional repressor activity.<sup>415</sup>

It is also well established that protein ubiquitination and SUMOylation are crucial in coordinating the cellular response to DSBs.<sup>416</sup> RNF4 is recruited to DNA lesions by damage-induced SUMOylation.<sup>417</sup> Here it facilitates the ubiquitination and dissociation of chromatin-associated MDC1 and RPA, thereby promoting NHEJ and HR DSB repair respectively.<sup>417</sup> It is believed that RNF4 contributes to DSB repair pathway choice in a cell cycle-dependent manner. It has been

shown that during S phase the ligase activity of RNF4 is enhanced by CDK2-mediated phosphorylation, inducing MDC1 degradation and therefore favouring HR repair.<sup>418</sup> Additionally, it was recently demonstrated that during G1 phase, RNF4 is degraded, permitting retention of DSB-induced phosphorylated Ser824 and SUMOylated TRIM28 (tripartite motif-containing 28), which blocks HR repair and promotes NHEJ.<sup>419</sup> When cells progress to S and G2, RNF4 accumulates and provokes pS824-TRIM28 degradation, relieving HR inhibition.<sup>419</sup> Further research indicates that ataxin-3, a deubiquitination enzyme, counteracts RNF4 activity during the DSB repair response, and that these opposing activities establish robust MDC1-dependent signalling and DSB repair.<sup>416</sup> It is proposed that ataxin-3 prevents MDC1 from being removed prematurely from the chromatin, to reinforce DNA damage signalling and properly initiate damage repair. Another study demonstrates that RNF4 is responsible for regulating the timely degradation of Fanconi anaemia protein complexes, as necessary during interstrand crosslink repair<sup>420</sup>, further attesting to the role of RNF4 in DSB repair.

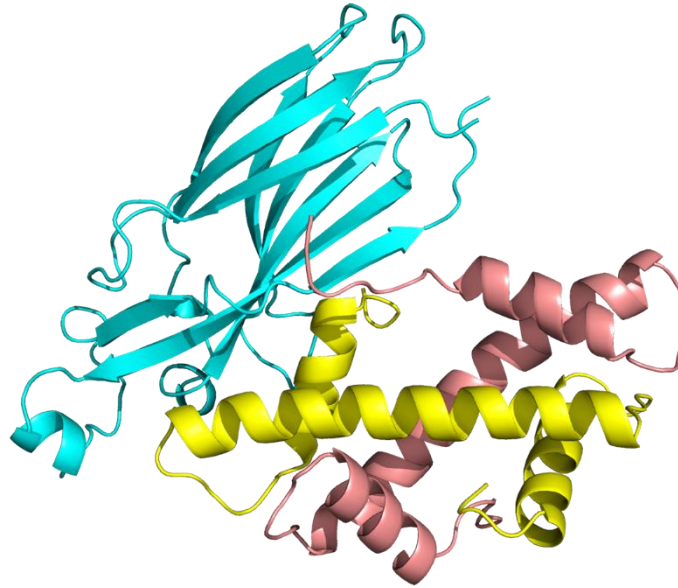


**Figure 4.3. Crystal Structure of RNF4.** Crystal structure of *Rattus norvegicus* RNF4 RING-type zinc finger domain homodimer (PDB ID: 3NG2)<sup>234</sup>, coloured from blue at N-terminus of chain A to red at C-terminus of chain B.

#### 4.1.4 ASF1A

ASF1A (anti-silencing function protein 1 homologue A) is an ATP-independent histone chaperone with anti-silencing activity<sup>421</sup>, capable of facilitating histone deposition, exchange and removal during nucleosome assembly or disassembly. It is specifically responsible for chaperoning histones H3 and H4 and has been characterised in complex with H3.1, and H4, Figure 4.4.<sup>236</sup>

The best understood role of ASF1A is in nucleosome assembly. It acts in cooperation with two other chaperones, CAF-1 (chromatin assembly factor 1) and HIRA (histone cell cycle regulator A), to mediate replication-coupled and replication-independent assembly respectively.<sup>422</sup> The interactions of ASF1A with the CAF-1 complex<sup>423</sup> and HIRA<sup>424</sup> are mutually exclusive. Nucleosome assembly and disassembly are essential in replication, transcription, and DNA repair processes and ASF1A has an extensive interaction network suggesting that it plays a critical role at the junction of chromatin and DNA checkpoint pathways.<sup>425</sup>



**Figure 4.4. Crystal Structure of ASF1A in Complex with Histones H3.1 and H4.** Crystal structure of *Homo sapiens* ASF1A (blue) in complex with histones H3.1 (yellow) and H4 (pink) (PDB ID: 2IO5).<sup>236</sup>

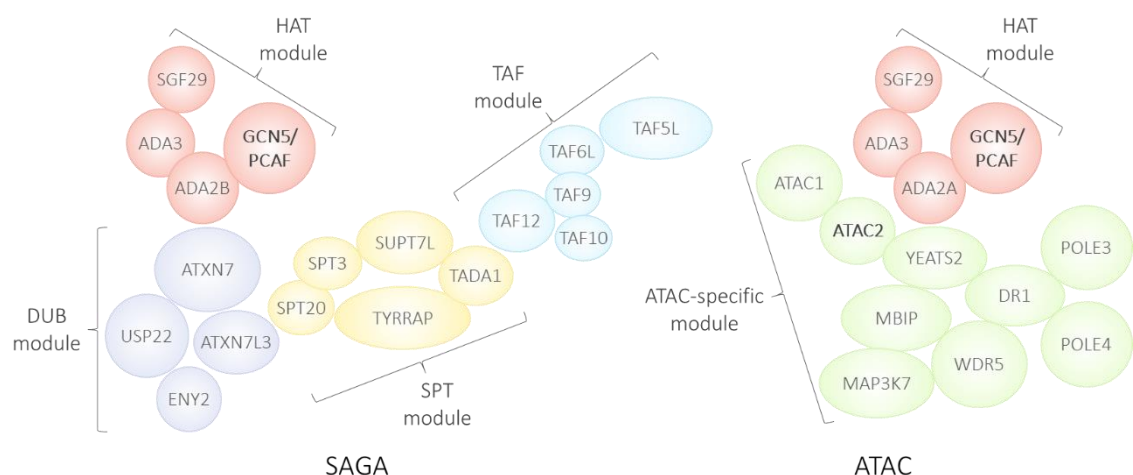
Correspondingly, ASF1A is implicated in a diverse array of cellular processes. As cells approach senescence, ASF1A and HIRA drive the formation of senescence-associated heterochromatin foci, at which proliferation-promoting genes are transcriptionally repressed, which provokes senescence-associated cell cycle exit.<sup>426</sup> Another interesting role has recently been determined in cellular reprogramming. ASF1A is required for H3K56 acetylation, which is associated with human embryonic stem cells, and overlaps with the binding of transcription factors key to regulating pluripotency, such as OCT4 (octamer-binding protein 4).<sup>427</sup> ASF1A is necessary for the reprogramming of human adult dermal fibroblasts into induced pluripotent stem cells (iPSCs).<sup>428</sup> Furthermore, another study demonstrated that p53 knockdown and ASF1A overexpression promoted the formation of iPSCs from sheep kidney cells.<sup>429</sup> ASF1A is also implicated in the DNA repair response. It has been shown that both nucleotide excision repair and mismatch repair occur during replication-coupled nucleosome assembly by CAF-1 and ASF1A.<sup>423,430</sup>

#### 4.1.5 GCN5/KAT2A

GCN5, or KAT2A (lysine acetyltransferase 2-A), is a histone acetyltransferase (HAT). These enzymes are responsible for acetylating conserved lysine residues within the N-terminal tails of histones by transferring acetyl groups from acetyl coenzyme A (CoA). Various studies published over the past few decades indicate that post-translational histone modifications serve to modulate chromatin compaction and consequently gene activity.<sup>381</sup> Specifically, actively transcribed DNA is often associated with acetylated histone isoforms, suggesting that HAT activity is correlated with transcriptional activation. This was first demonstrated by Hebbes *et al.* in 1988.<sup>431</sup> Antibodies raised against hyperacetylated histones were employed in chromatin

immunoprecipitation assays to illustrate that these histones tend to localise to regions which are transcriptionally active.<sup>431</sup> Histone acetylation is now established as a hallmark of transcriptionally competent chromatin, alongside sensitivity to nuclease digestion.<sup>432</sup> It is believed that acetylation activates gene expression by loosening the chromatin structure, due to charge neutralisation of the basic lysines. This weakens the interaction of the histone with the DNA, and establishes docking sites for other proteins and complexes.<sup>433</sup> In contrast to that of SWI/SNF complexes, this chromatin remodelling function is ATP-independent.<sup>434</sup>

Vertebrates have another HAT called PCAF (P300/CREBBP-associated factor), or KAT2B. KAT2A and KAT2B have a sequence identity of 70%.<sup>435</sup> Neither exist freely in cells. The two homologues are integrated mutually exclusively into one of two related complexes, SAGA (Spt-Ada-Gcn5 acetyltransferase) or ATAC (ADA2A-containing).<sup>436</sup> Until relatively recently there were believed to be three unique complexes, STAGA (SPT3-TAF9-GCN5 acetyltransferase), TFTC (TBP-free-TAF-containing) and ATAC. However, it has been concluded that STAGA and TFTC are the same, and the original preparations can be separated into SAGA and a series of TAF (TBP-associated factor)-containing complexes.<sup>437</sup> Yeast have a single SAGA complex, which has an assortment of functions, such as in transcription initiation and elongation, TBP interactions, mRNA export and maintaining DNA integrity. This complex has diverged during evolution to produce the four SAGA-like complexes in humans, SAGA and ATAC, each with two variants, incorporating KAT2A or KAT2B<sup>438</sup>, Figure 4.5. SAGA is the larger of the multiprotein complexes, composed of 18 subunits arranged into HAT, TAF, DUB (deubiquitination) and SPT (suppressor of Ty) modules. The physical relationships between the subunits in ATAC are less well defined, but it is clear that the HAT module is retained, with ADA2A replacing ADA2B, and the DUB module is not.<sup>436</sup> Interestingly, ATAC incorporates two HAT subunits, GCN5/PCAF and ATAC2.<sup>438</sup>



**Figure 4.5. Schematic of Mammalian SAGA-Like Complexes.** Structures of the multiprotein SAGA and ATAC complexes, summarising the modules and subunits incorporated. Subunits with HAT activity are labelled in bold. Figure adapted from Wang & Dent, 2014.<sup>436</sup>

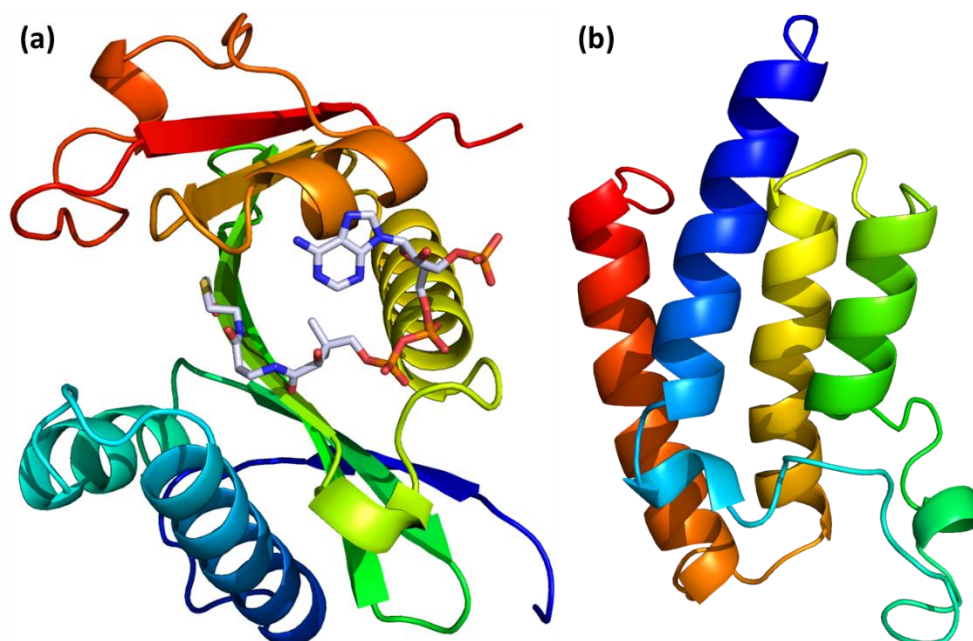
The enzymatic activity of GCN5 is enhanced by the other subunits within the complexes.<sup>435</sup> SAGA and ATAC exhibit HAT activity at histones H3 and H4. However, it appears that GCN5 preferentially acetylates H3K9 and H3K14 residues. Correspondingly, it has been demonstrated that RNAi depletion of GCN5 results in a decrease in the acetylation of H3 but not H4.<sup>439</sup> It is likely that several HATs act redundantly at H4. The specificity of ATAC2 is less well understood, but it appears to show some preference for H4K16.<sup>440</sup> The HAT activities of both GCN5 and ATAC2 contribute to the specificity of ATAC. SAGA also exhibits a second enzymatic role in deubiquitination, via the USP22 (ubiquitin-specific protease 22) subunit. The complex has been identified as the major H2B deubiquitinase in human cells and this deubiquitination activity is required for full activation of SAGA-dependent inducible genes.<sup>441</sup>

SAGA and ATAC have a multitude of functions, for example in stress-induced signalling pathways, chromatin remodelling and transcription regulation, embryonic development, cell cycle progression and DNA repair.<sup>438</sup> Principally, SAGA is critical in transcription initiation. By acetylating histones at promoter regions and deubiquitinating those within transcribed regions, the complex recruits RNA polymerase II and thereby nucleates transcription initiation. It appears that SAGA is the cofactor for all RNA polymerase II transcription.<sup>442</sup> The ATAC complex is particularly crucial for mammalian development. *Atac2*-null mice display early embryonic lethality due to cell cycle defects.<sup>443</sup>

Interestingly, in human immunodeficiency virus type 1 (HIV-1) infection, KAT2A is known to specifically interact with and acetylate the viral transactivator protein, Tat, which enhances Tat-dependent transcription of the HIV-1 long terminal repeat.<sup>444</sup> It is also responsible for the acetylation of HIV-1 integrase, which catalyses the integration of the viral DNA into the host genome. KAT2A knockdown cells demonstrate reduced infectivity.<sup>445</sup>

KAT2A contains the catalytic HAT domain and a bromodomain, illustrated in Figure 4.6 (a)<sup>241</sup> and (b)<sup>238</sup> respectively. As discussed, HATs acetylate conserved lysine residues in the amino terminal domains of the histones, to induce transcriptional activity.<sup>381</sup> The HAT domain is the catalytic domain responsible for transferring acetyl groups from acetyl-CoA to the lysine residues. Bromodomains are able to recognise and selectively bind these acetylated lysine residues. Perhaps predictably, these evolutionarily conserved protein modules are found in many chromatin-associated proteins and the majority of known nuclear HATs. It is thought that bromodomains facilitate highly specific histone acetylation by tethering the HATs to precise chromosomal sites. In three-dimensional structure the domain adopts a left-handed four-helix bundle, with the two intervening loops at one end packing to form a surface-accessible hydrophobic pocket for binding at acetylated lysine.<sup>446</sup>

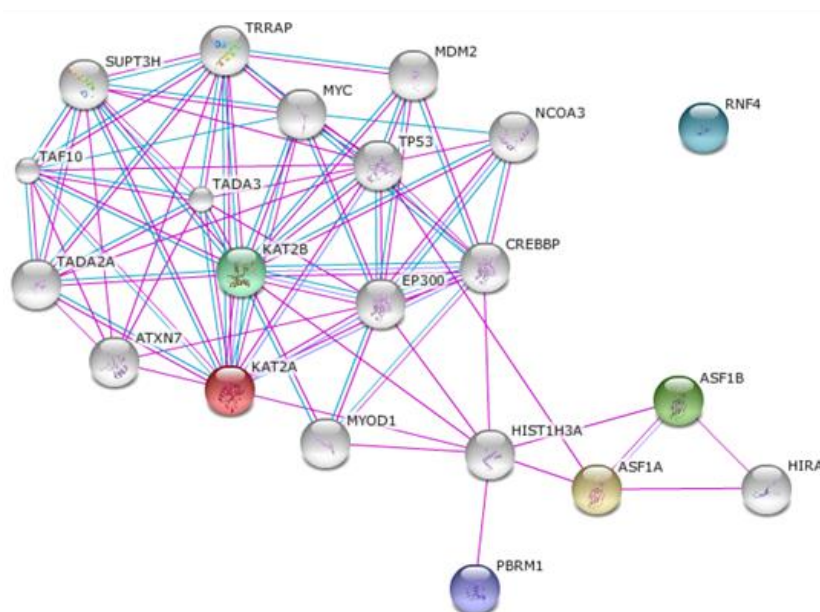




**Figure 4.6. Crystal Structures of GCN5.** (a) Crystal structure of *Homo sapiens* GCN5 acetyltransferase domain with acetyl-CoA (grey) bound (PDB ID: 1Z4R).<sup>241</sup> (b) Crystal structure of *Homo sapiens* GCN5 bromodomain (PDB ID: 3D7C, chain A).<sup>238</sup> Protein chains coloured from blue at N-terminus to red at C-terminus.

## 4.2 Results

Having selected RNF4, ASF1A and GCN5 as biologically desirable drug targets, information and data on these proteins was compiled from online resources and databases with an aim to assess their druggability.



**Figure 4.7. Schematic of Functional Protein Association Networks.** Association networks of PBRM1 and RNF4, ASF1A and KAT2A, predicted using results from experiments (pink) and databases (blue), determined by STRING 9.1.<sup>161</sup>

A diagram of the relevant protein association networks was constructed in STRING 9.1<sup>161</sup>, using results from experiments and databases to predict association and incorporating 15 additional white nodes. The resulting schematic is shown in Figure 4.7. While GCN5 homologues, KAT2A

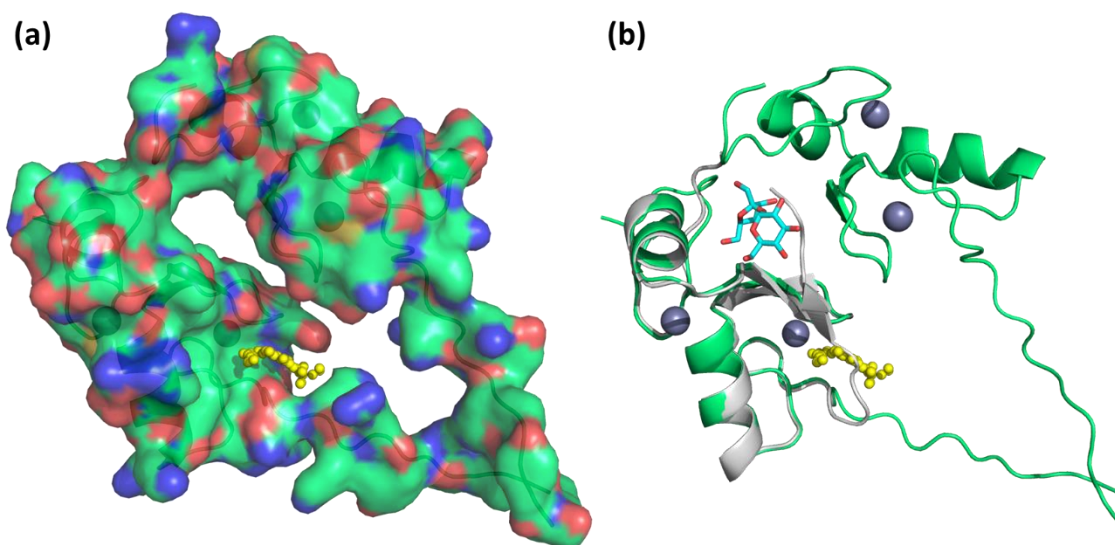


and KAT2B appear to interact with an array of proteins, ASF1A was shown only to interact with the histones, HIRA, p53 and ASF1B, and no interacting partners of RNF4 had yet been identified. This may be indicative of the number of cellular processes the proteins are involved in.

#### 4.2.1 RNF4

Due to the small size of RNF4, it was suggested that it should be relatively easy to produce to use in assays. Furthermore, the low number of cellular processes in which it is implicated might reduce the number of on-target side effects of a targeted therapeutic. From a biological viewpoint, RNF4 was therefore an attractive target.

A DELTA-BLAST<sup>180,183</sup> search against the PDB identified a crystal structure of the *Rattus norvegicus* RNF4 (PDB ID: 3NG2)<sup>234</sup>, which showed 99% identity with the human orthologue at the RING finger domain region, and therefore no modelling was required. Binding pockets were identified from receptor cavities using Fpocket 2.0<sup>165</sup>, and the corresponding Fpocket druggability scores were calculated.<sup>170</sup> Only one pocket was identified, as highlighted in Figure 4.8 (a), with an associated druggability score of 0.608, which is higher than the threshold of 0.5, suggesting that the cavity may be druggable.



**Figure 4.8. Analysis of Structural Pocket in RNF4, Identified by Fpocket 2.0.**<sup>165</sup> (a) *Rattus norvegicus* RNF4 homodimer (PDB ID: 3NG2)<sup>234</sup> shown in green with heteroatoms highlighted. Structural pocket depicted in yellow. (b) *Homo sapiens* RNF4 (PDB ID: 2XEU)<sup>235</sup> in grey, aligned with that of *Rattus norvegicus* in green. Structural pocket shown in yellow and ubiquitin moiety in cyan. The interaction site of the ubiquitin moiety does not align with the identified cavity.

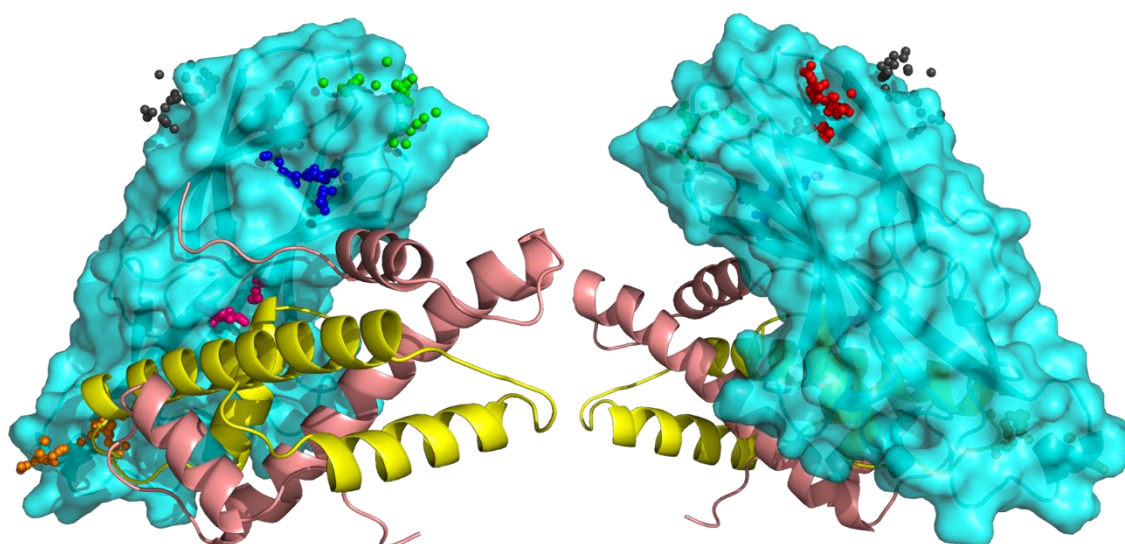
The PDB<sup>158</sup> and ChEMBL<sup>159</sup> were searched to identify small molecules which could bind at this site and thus provide inspiration for a pharmacophore. Unfortunately, this search did not yield any positive results. A crystal structure of *Homo sapiens* RNF4 (PDB ID: 2XEU)<sup>235</sup> appeared to depict a small molecule binding to RNF4. However, upon closer inspection it became apparent that the ligand in question was actually the 'Ile44 hydrophobic patch' of the larger ubiquitin

protein. Furthermore, alignment of 2XEU<sup>235</sup> with 3NG2<sup>234</sup> confirmed that the site of this interaction did not correspond with the position of the cavity identified by Fpocket, Figure 4.8 (b). In summary, with the information available, RNF4 did not represent a promising drug target.

#### 4.2.2 ASF1A

ASF1A is larger than RNF4 and it may therefore be more difficult to produce stable protein, however, similarly to RNF4, the low number of cellular processes in which it has been implicated might reduce the potential for on-target side effects.

A DELTA-BLAST<sup>180,183</sup> search against the PDB identified the *Homo sapiens* crystal structure of ASF1A complexed with histones H3 and H4 (PDB ID: 2IO5)<sup>236</sup>, and therefore no modelling was necessary. Fpocket 2.0<sup>165</sup> analysis identified six receptor cavities, which are displayed in Figure 4.9, and the associated Fpocket druggability scores<sup>170</sup> are shown in Table 4.1. While five were very low scoring, one (shown in red) was very interesting, having an associated druggability score of 0.963, suggesting a highly druggable cavity.



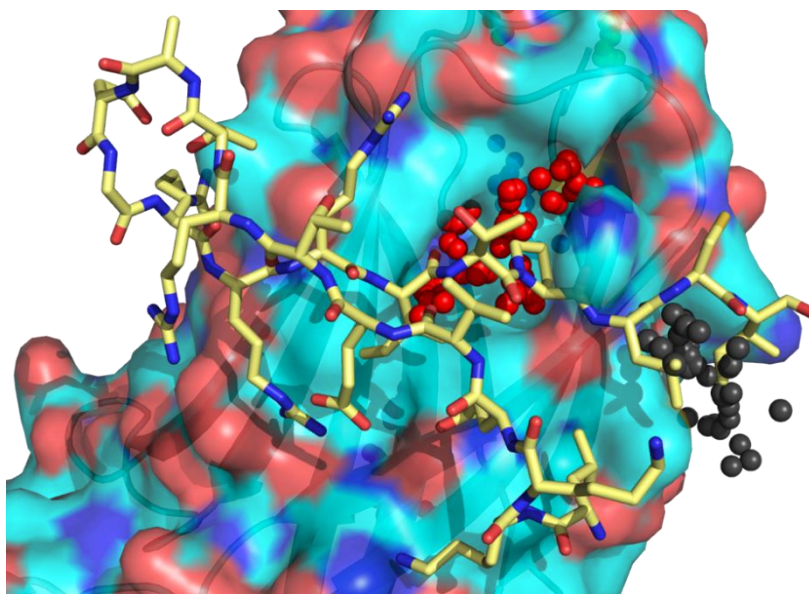
**Figure 4.9. Analysis of Structural Pockets of ASF1A Identified by Fpocket 2.0.**<sup>165</sup> Rotated views of *Homo sapiens* ASF1A complexed with histones H3 and H4 (PDB ID: 2IO5)<sup>236</sup>, with ASF1A shown in cyan, histone H3 and H4 in yellow and pink respectively and structural pockets highlighted.

Surprisingly, the high scoring pocket was not located at the PPI interface with histones H3.1 and H4, but on the opposite face of ASF1A. Upon investigation, it was found that this pocket was in fact the interaction site for HIRA. By comparing 2IO5<sup>236</sup> with a crystal structure of the *Homo sapiens* ASF1A-HIRA complex (PDB ID: 2I32)<sup>424</sup>, Figure 4.10, it was found that the cavity coincided with the binding position of a proline residue at the ASF1A-HIRA PPI interface. This ASF1A-HIRA complex had been identified during an interesting study which used targeted mutagenesis as a means to biochemically characterise the interaction between HIRA and ASF1A.<sup>424</sup>

**Table 4.1. Druggability Scores from Fpocket 2.0.**<sup>170</sup> Druggability scores for the pockets identified in ASF1A.

Pocket	Colour	Druggability score
1	Red	0.963
2	Green	0.020
3	Pink	0.062
4	Orange	0.026
5	Blue	0.086
6	Grey	0.011

While this structural cavity may be druggable, this is of little use if such binding does not induce the desired biological effect. The protein association network generated in STRING 9.1<sup>161</sup> suggests that PBRM1/BAF180 and ASF1A are likely to be synthetic lethal partners by virtue of their mutual interaction with the histones. Consequently, a small molecule which targets the ASF1A-HIRA PPI interface may not confer the necessary biological response to result in cell death in combination with a mutation in BAF180. Unfortunately, in unpublished work (Hopkins S. and Downs J., Genome Damage and Stability Centre, University of Sussex), it was confirmed that HIRA and BAF180 are not synthetic lethal, suggesting that a drug targeted at the ASF1A-HIRA PPI interface is unlikely to enable selective killing of cancer cells mutated in BAF180. Given the absence of other druggable pockets, it was concluded that ASF1A did not represent a worthwhile target for tool inhibitor development at this time.



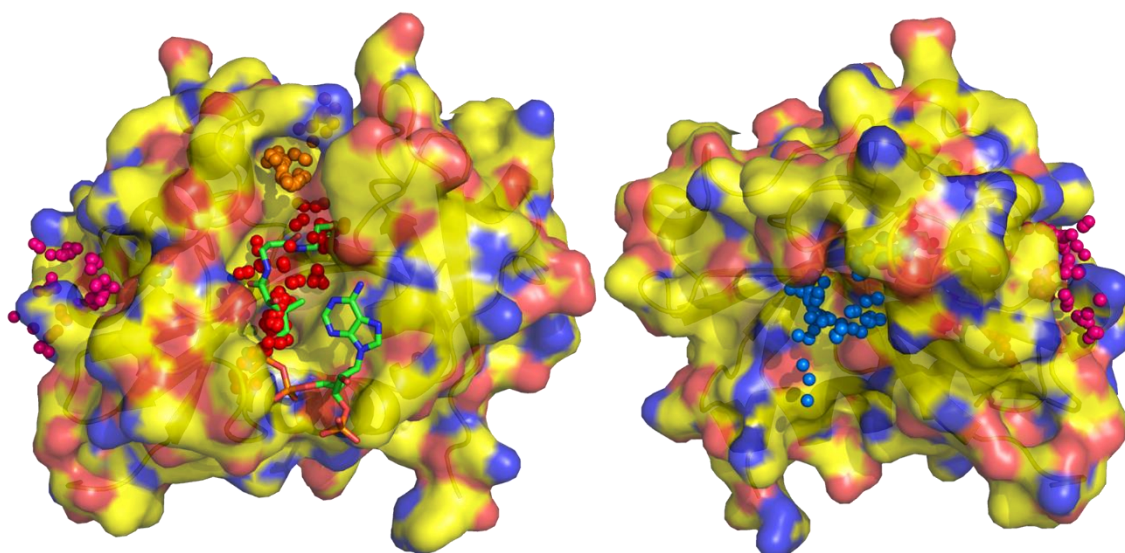
**Figure 4.10. Comparing the PPI of ASF1A and HIRA with Structural Pockets Identified by Fpocket 2.0.**<sup>165</sup> *Homo sapiens* ASF1A (PDB ID: 2IO5)<sup>236</sup> aligned with *Homo sapiens* ASF1A-HIRA complex (PDB ID: 2I32).<sup>424</sup> ASF1A is shown in cyan and HIRA in cream, with the heteroatoms highlighted. Structural pockets are displayed as spheres. Red pocket corresponds to binding position of proline residue.

#### 4.2.3 GCN5/KAT2A

KAT2A is a HAT, responsible for promoting transcriptional activation.<sup>447</sup> As HATs are well studied, many suitable and relatively simple assays already exist which could be utilised in screening.

However, unlike RNF4 and ASF1A, the high number of cellular processes in which the protein is implicated may increase the potential for adverse effects from targeted therapy.

A DELTA-BLAST<sup>180,183</sup> search against the PDB yielded two discrete classes of matches, showing sequence similarity at either the HAT domain or the bromodomain. The crystal structures of the *Homo sapiens* GCN5 Acetyltransferase domain (PDB ID: 1Z4R)<sup>241</sup>, and the *Homo sapiens* GCN5 bromodomain (PDB ID: 3D7C)<sup>238</sup>, were identified and therefore again no modelling was needed. The binding pockets identified from receptor cavities using Fpocket 2.0<sup>165</sup> are shown in Figure 4.11 and Figure 4.12 respectively, and the corresponding druggability scores<sup>170</sup> in Table 4.2. Fpocket identified four pockets at the HAT domain, one of which exhibited a particularly high druggability score of 0.908, and another four at the bromodomain, two of which had associated druggability scores of 0.807 and 0.869, suggestive of highly druggable cavities.



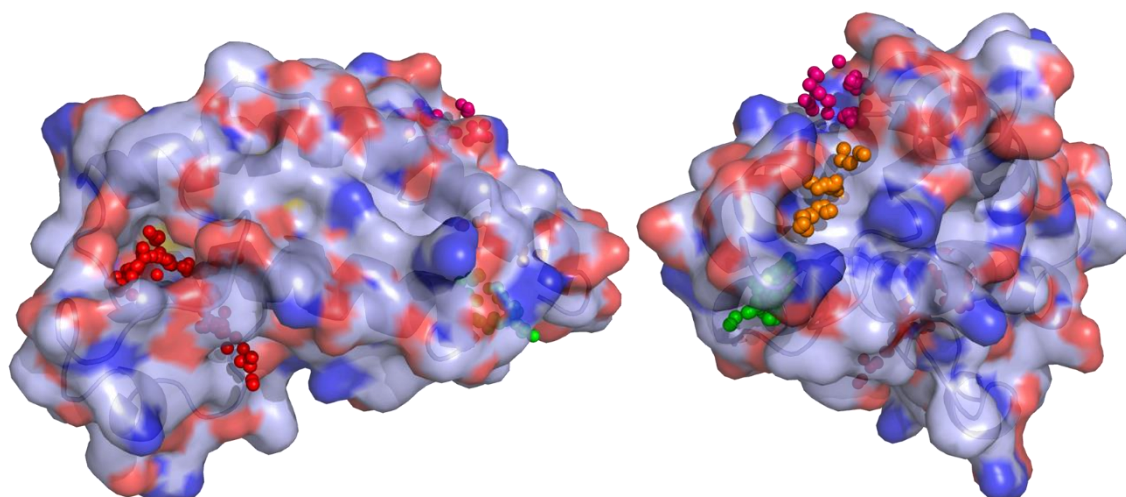
**Figure 4.11. Structural Pockets of KAT2A Acetyltransferase Domain Identified by Fpocket 2.0.**<sup>165</sup> Rotated views of *Homo sapiens* GCN5 acetyltransferase domain (PDB ID: 1Z4R)<sup>241</sup>, with KAT2A in yellow and acetyl-CoA in green, with heteroatoms highlighted. Structural pockets shown as spheres. Red pocket corresponds to binding site of acetyl-CoA.

Unsurprisingly the most druggable cavity identified at the HAT domain was found to coincide with the binding site of acetyl-CoA. This is useful as the small molecule can be utilised in designing potential pharmacophores for drug development, and a tool inhibitor successfully targeted at this site would doubtlessly disrupt enzymatic function. However, given that all HATs are able to bind acetyl-CoA, selectively targeting this site may be challenging.

3D7C<sup>238</sup> is an apo structure, depicting the unbound KAT2A bromodomain. However, as discussed above, bromodomains are known to interact with acetylated lysine residues. To ascertain whether the lysine binding site coincided with either of the pockets identified as druggable, 3D7C<sup>238</sup> was aligned with a crystal structure of *S. cerevisiae* GCN5 bromodomain complexed with H4 peptide acetylated at lysine 16 (PDB ID: 1E6I)<sup>239</sup>, Figure 4.13. It was discovered that the



acetylated peptide binds at the site that has an associated druggability score of 0.807. As above, the structure of the acetylated peptide ligand is useful in suggesting potential pharmacophores for use in tool inhibitor design. It also improves the likelihood that drugs targeted at this site should have a desirable therapeutic effect. It is proposed that preventing the bromodomain from successfully tethering the HAT at precise chromosomal sites should prevent the proper acetylation of histones.

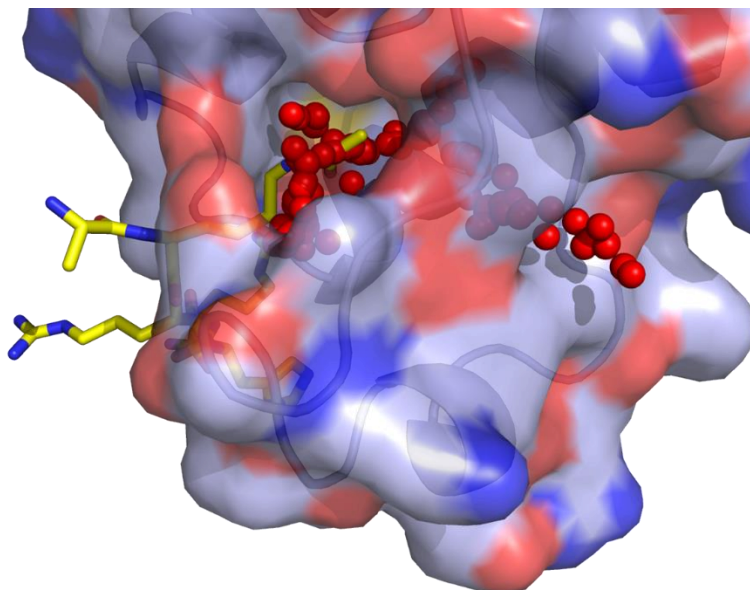


**Figure 4.12. Structural Pockets of KAT2A Bromodomain Identified by Fpocket 2.0.**<sup>165</sup> Rotated views of *Homo sapiens* GCN5 bromodomain (PDB ID: 3D7C)<sup>238</sup>, with KAT2A shown in purple with the heteroatoms highlighted. Structural pockets shown as spheres.

**Table 4.2. Druggability Scores from Fpocket 2.0.**<sup>170</sup> Druggability scores for the pockets identified at the KAT2A HAT domain and bromodomain.

	Pocket	Colour	Druggability Score
KAT2A Acetyltransferase Domain	1	Red	0.908
	2	Blue	0.171
	3	Pink	0.043
	4	Orange	0.045
KAT2A Bromodomain	1	Red	0.807
	2	Green	0.869
	3	Pink	0.031
	4	Orange	0.038

Given the high similarity of the two mammalian GCN5 homologues, KAT2A and KAT2B, both in their structures and involvement in protein association networks, it was deemed important to test the effect of knocking down KAT2B in BAF180 deficient cells. Interestingly, in unpublished work (Hopkins, S. and Downs, J., Genome Damage and Stability Centre, University of Sussex), it was determined that KAT2B is not synthetically lethal with PBRM1, highlighting KAT2A as a particularly valuable target for tool inhibitor development. Despite inevitable challenges, from the information available, KAT2A therefore appears to be an interesting and promising target for tool inhibitor development at both the HAT domain and bromodomain.



**Figure 4.13. Analysis of Druggable Pocket in KAT2A Bromodomain.** *Homo sapiens* GCN5 bromodomain (PDB ID: 3D7C)<sup>238</sup>, shown in purple with the heteroatoms highlighted, aligned with that of *S. cerevisiae* (PDB ID: 1E6I)<sup>239</sup>, to determine whether the binding site of acetylated lysine corresponds to one of the structural pockets identified. Histone H4 peptide acetylated at lysine 16 in yellow with the heteroatoms highlighted and structural cavity in red.

### 4.3 Discussion

Mutations in PBRM1 have been identified in a variety of human cancers<sup>377</sup>, but most recently the gene has been identified as a major player in clear cell renal cell carcinoma, with truncating mutations observed in 41% of samples of primary ccRCC.<sup>1</sup> Such unique features of tumours can be exploited in targeted therapies and are currently the focus of many efforts in cancer research using synthetic lethality approaches. The prominent mutation of PBRM1 has been specifically highlighted as a promising opportunity to target ccRCC.<sup>2</sup> The aim of this work was to assess the druggabilities of RNF4, ASF1A and GCN5, which have been identified as synthetic lethal partners of PBRM1<sup>3</sup>, and therefore could represent interesting targets for tool inhibitor development. Such tools should improve biological investigation and could initiate drug discovery projects.

A crystal structure of *Rattus norvegicus* RNF4 (PDB ID: 3NG2)<sup>234</sup> was identified with 99% identity to the human orthologue. A search for structural pockets identified one of particular interest with an associated druggability score of 0.608, suggesting it would be druggable. However, a search for small molecules bound at this site, which could provide inspiration for a corresponding pharmacophore, was unsuccessful. With the information available, RNF4 did not therefore represent a promising drug target.

A crystal structure was identified characterising *Homo sapiens* ASF1A complexed with histones H3 and H4 (PDB ID: 2IO5).<sup>236</sup> One interesting pocket was identified with an associated druggability score of 0.963. This pocket was located not at the PPI interface with histones H3.1 and H4, but at the ASF1A-HIRA interface. It was noted that inhibiting ASF1A and HIRA from

interacting may not confer the desired synthetic lethal response with PBRM1 mutation. In unpublished work (Hopkins S. and Downs J., Genome Damage and Stability Centre, University of Sussex), it was confirmed that HIRA and PBRM1 are not synthetic lethal, suggesting that a drug targeted at the ASF1A-HIRA interface is not likely to induce selective killing in combination with PBRM1 mutation. Given the absence of other druggable pockets, it was determined that ASF1A was not a promising drug target at this time.

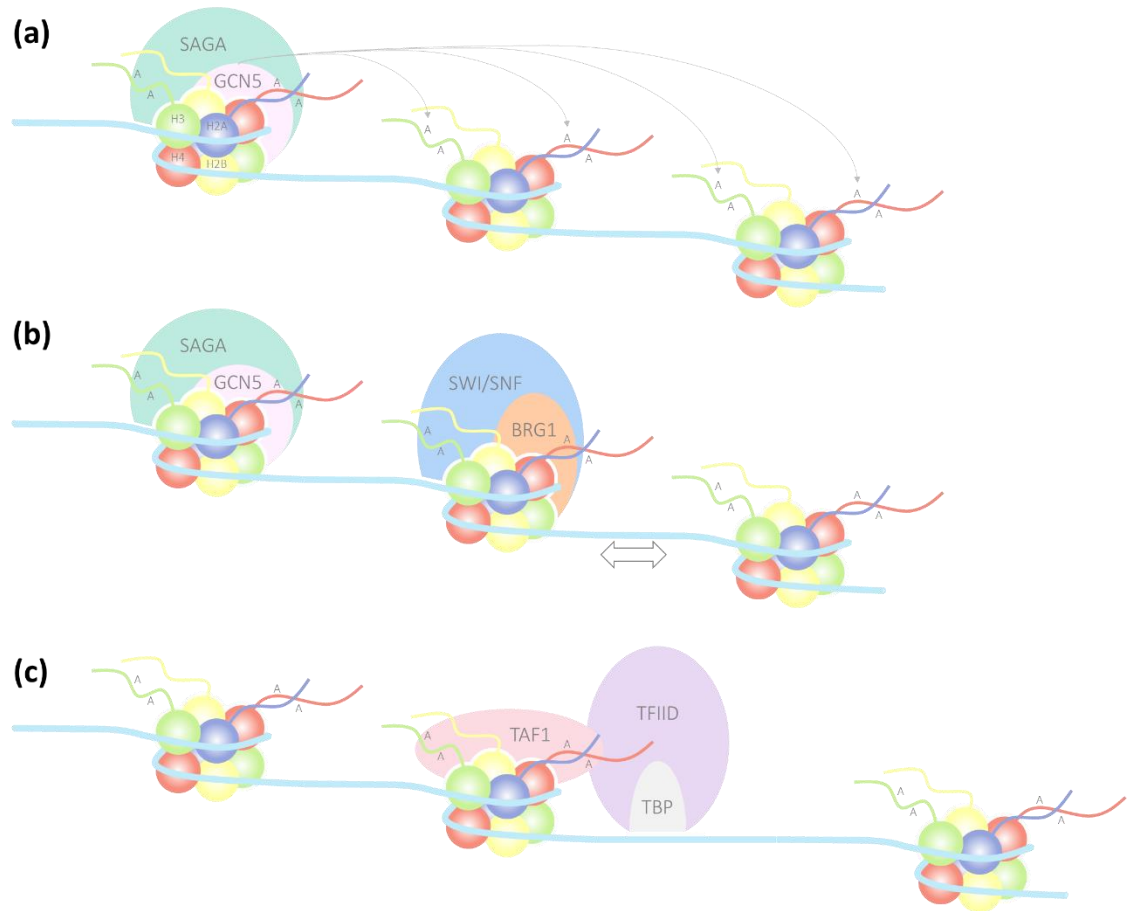
KAT2A contains both a HAT domain and a bromodomain. Crystal structures of the *Homo sapiens* GCN5 Acetyltransferase domain (PDB ID: 1Z4R)<sup>241</sup>, and the *Homo sapiens* GCN5 bromodomain, (PDB ID: 3D7C)<sup>238</sup>, were identified. One pocket at the HAT domain, with a particularly high druggability score of 0.908, was found to coincide with the binding site of acetyl-CoA, thereby suggesting potentially suitable pharmacophores which could be utilised in drug design. A pocket of interest was also identified at the bromodomain, which had an associated druggability score of 0.807. This pocket coincided with the binding site of acetylated lysine residues, which again should be useful in tool inhibitor design, towards identifying pharmacophores and developing assays. Because mammalian cells have two GCN5 homologues, KAT2A and KAT2B, which are implicated in similar pathways, it seemed important to test whether KAT2B knockdown effects BAF180 deficient cells. This was investigated in unpublished work (Hopkins S. and Downs J., Genome Damage and Stability Centre, University of Sussex), and it was confirmed that KAT2B and PBRM1 are not synthetic lethal. From the available information, KAT2A was established as a promising and interesting target for tool inhibitor development.

In summary, at the time this work was completed, RNF4 and ASF1A did not represent druggable targets, within the restraints of accessible resources. They could be revisited in the future, if further data and biological resources become available. However, KAT2A was highlighted as a promising target for development of a tool inhibitor.

#### 4.3.1 KAT2A-BAF180 Synthetic Lethality

The precise mechanism for KAT2A-BAF180 synthetic lethality has not been defined. However, the two proteins are engaged in a number of overlapping pathways. The relationship between histone acetylation and chromatin remodelling was first observed in yeast. Genetic experiments were used to demonstrate that mutations in GCN5 and SWI/SNF cause strikingly similar phenotypes; both are required for proper expression of the same genes, interact with the same chromatin components, and are synthetic lethal in yeast.<sup>448</sup> Furthermore, chromatin immunoprecipitation studies indicate that GCN5 is involved in recruiting and maintaining SWI/SNF at promoters<sup>449,450</sup>, and that in eukaryotes histone acetylation occurs prior to SWI/SNF

recruitment.<sup>451,452</sup> Finally, it has been observed that GCN5-dependent acetylation at H3K9, H3K14, H4K8 and H4K12 is required for recruitment of the SWI/SNF BRG1 subunit, via acetylated H4K8<sup>453</sup>, and transcription factor II D (TFIID) via TAF1, which interacts with all four acetylated lysines<sup>453–455</sup>, Figure 4.14.



**Figure 4.14. Examples of Mechanisms by which GCN5 Regulates Gene Expression.** (a) The GCN5 bromodomain binds to acetylated histones and anchors SAGA at the nucleosome, enabling acetylation of neighbouring nucleosomes. (b) BRG1 subunit of SWI/SNF binds to acetylated H4K8, enabling SWI/SNF to remodel adjacent promoter region. (c) TAF1 binds to acetylated nucleosome and recruits transcription factor TFIID. Figure adapted from Josling *et al.*, 2012.<sup>456</sup>

Similarly, histone acetylation by the SAGA complex has been observed to promote SWI/SNF recruitment *in vitro*. SWI/SNF exhibits a higher affinity for nucleosomes containing acetylated H3, such that it is retained at these nucleosomes even after dissociation of the participating transcriptional activator.<sup>457</sup> This could suggest that recruitment by lysine acetylation occurs by slowing the off-rate rather than promoting initial association.<sup>434</sup> In support of this hypothesis, removing the bromodomain from the BRG1-like subunit disrupts but does not abolish recruitment<sup>458</sup>, which is less severe than the effect of mutations obstructing interactions with transcriptional activators.<sup>459</sup> These observations suggest that chromatin remodellers are recruited to regulatory regions via site-specific DNA binding factors and then interact via their bromodomains with acetylated histones to stabilise their association with the chromatin.<sup>434</sup>



Mutations affecting RSC, another SWI/SNF homologue, are also lethal in combination with GCN5 knockout, and this complex interacts with acetylated H3 as well.<sup>460</sup> GCN5 is actually implicated in both positive and negative regulation of RSC. It has been observed that GCN5 acetylates RSC4 K45 to create an intramolecular binding site for one of the RSC4 bromodomains. The associated binding prevents a second RSC4 bromodomain from interacting with H3K14.<sup>461</sup> Despite the strong influence on binding affinity, it is rare that histone acetylation dramatically impacts the ATPase or remodelling functional activities of SWI/SNF complexes.<sup>462</sup>

GCN5-dependent histone acetylation has also been shown to recruit SWI/SNF complexes to DNA double-stranded breaks via the BRG1 bromodomain.<sup>46</sup> In mammalian cells, recruitment of BAF is reliant on  $\gamma$ H2AX, but these do not appear to interact directly.<sup>463</sup> Lee *et al.* showed that  $\gamma$ H2AX recruits GCN5, which acetylates H3 to promote BAF recruitment.<sup>46</sup> The chromatin remodelling directed by BAF increases the accessibility of neighbouring nucleosomes, enabling further H2AX phosphorylation and thereby more acetylation by GCN5 and further increased BAF binding. This results in a cooperative activation loop to propagate  $\gamma$ H2AX formation and induce chromatin remodelling at DSBs.<sup>46</sup> The interaction of SAGA with promoters seems to depend on a similar feed forward loop regulated by histone modifications. SAGA interacts with acetylated H3 via the GCN5 bromodomain.<sup>458</sup> Thus, it appears that SAGA writes the acetylation mark on histone H3, and subsequently reads it. In this way GCN5 could be proposed to walk along a length of chromatin, gradually acetylating histones so as to establish a transcriptionally competent structure.<sup>464</sup>

In conclusion, while the mechanism for KAT2A-BAF180 synthetic lethality is not yet understood, it is clear that KAT2A has the potential to be an interesting drug target. To understand whether the druggability determined in this work translates to phenotypic activity against BAF180 deficient cancer cells, KAT2A tool inhibitors are required. Work carried out towards developing tool inhibitors for the KAT2A HAT domain and bromodomain is described in Chapters 5 and 6 respectively. It is hoped that these tools will enable detailed biological investigation to afford a better understanding of the dynamic regulation of gene expression by acetylation, and potentially initiate translational drug discovery projects towards treating ccRCC via a synthetic lethality approach.

## 5 Towards Development of KAT2A HAT Domain Tool Inhibitors

KAT2A is synthetic lethal with BAF180<sup>3</sup>, the second most frequently mutated gene in clear cell renal cell carcinoma.<sup>376</sup> This histone acetyltransferase is therefore an interesting target for tool inhibitor development, to enable biological investigation and potentially initiate translational drug discovery projects towards treating ccRCC.

In Chapter 4.2, a highly druggable binding pocket was identified at the KAT2A HAT domain using computational techniques. HATs represent an exciting class of targets in drug discovery, with an array of potential applications in many diseases. They are commonly deemed interesting targets for development of tool inhibitors, even in the absence of specific disease association. However, existing KAT2A inhibitors are inadequate for biological investigation, disease validation and initiation of drug discovery efforts.

In this chapter, computational techniques were used to determine that a selective dual inhibitor of the KAT2A and KAT2B HAT domains was achievable and to select a diverse set of compounds with reported activity against KAT2A. Unfortunately, attempts to establish a fluorescence based activity assay in house, which was previously employed in high-throughput screening, were unsuccessful as the assay proved prohibitively unreliable. With those hits from this screen discounted<sup>4,5</sup>, the inventory of small molecule inhibitors of KAT2A and KAT2B was quickly exhausted, thereby terminating attempts to develop a KAT2A HAT tool inhibitor.

### 5.1 Introduction

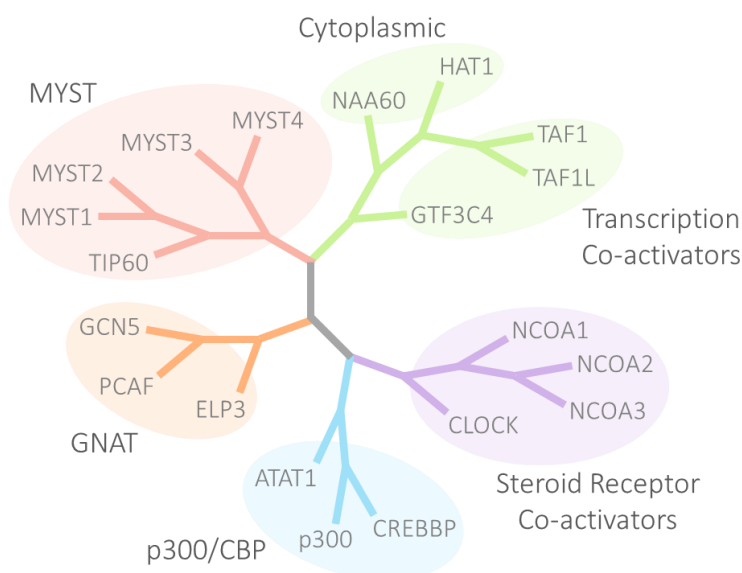
Based on the previous work, in which KAT2A was identified as a promising target for development of a tool inhibitor, this enzyme became the major focus of this research. As discussed, KAT2A contains a HAT domain and a bromodomain. The HAT domain is the catalytic domain responsible for transferring acetyl groups from acetyl-CoA to conserved lysine residues in the amino terminal domains of histones, to induce transcriptional activity.<sup>381</sup> An inhibitor successfully designed to target this domain and disrupt the interaction with acetyl-CoA would therefore disrupt enzymatic function and afford the desired biological response. As a result, the HAT domain was prioritised as the preferred target for tool inhibitor development.

#### 5.1.1 *Phylogeny, Structure and Function*

Protein acetylation was discovered in 1963, when Phillips identified acetyl groups in histones isolated from calf thymus.<sup>465</sup> In 1964, Allfrey *et al.* showed that in isolated nuclei radiolabelled acetate was rapidly incorporated into histones, independent of translation.<sup>466</sup> Furthermore, it had not long before been shown that histones inhibit RNA synthesis, and Allfrey *et al.*

demonstrated that histone acetylation decreased the effectiveness of this inhibition. They proposed that post-translational histone acetylation affords a 'dynamic and reversible mechanism for "activation" as well as "repression" of RNA synthesis'.<sup>466</sup> Unfortunately, after this development, acetylation was largely neglected from research efforts for 30 years<sup>467</sup>, with the notable exception that, in 1978, sodium butyrate was discovered to inhibit histone deacetylase activity, causing histone hyperacetylation.<sup>468,469</sup> In 1988, the connection between histone acetylation and transcriptional activation was uncovered<sup>431</sup> and there was growing recognition that histones acetylated at specific residues mediate unique effects on gene expression.<sup>467</sup> This initiated a flood of research in the 1990s, affording many discoveries, including the pivotal revelations that GCN5 and RPD3 (reduced potassium dependency 3), both known transcription regulators, possess histone acetyltransferase<sup>470</sup> and deacetylase<sup>471</sup> activity respectively. This finally confirmed a causal link between histone acetylation and transcriptional regulation.

It is now widely accepted that transcription can be regulated via acetylation of conserved lysine residues located in the amino terminal domains of histones. Acetylation results in charge neutralisation which is thought to weaken the interactions between the histones, DNA and regulatory proteins, which alters the nucleosome structure, producing the more open chromatin environment required for transcription.<sup>381</sup> These acetylation marks are written by HATs, by transferral of an acetyl from acetyl-CoA, and erased by histone deacetylases (HDACs).



**Figure 5.1. Phylogenetic Tree of HAT Families.** Schematic showing HATs subdivided into their respective families. Figure adapted from Arrowsmith *et al.*, 2012.<sup>380</sup>

During the following five years an array of additional HATs were identified<sup>467</sup>, including for example, TAF1<sup>472</sup>, p300 and CREBBP<sup>473</sup>. Phylogenetic trees have been devised by groups such as the Structural Genomics Consortium (SGC) to map the evolutionary and structural relationships

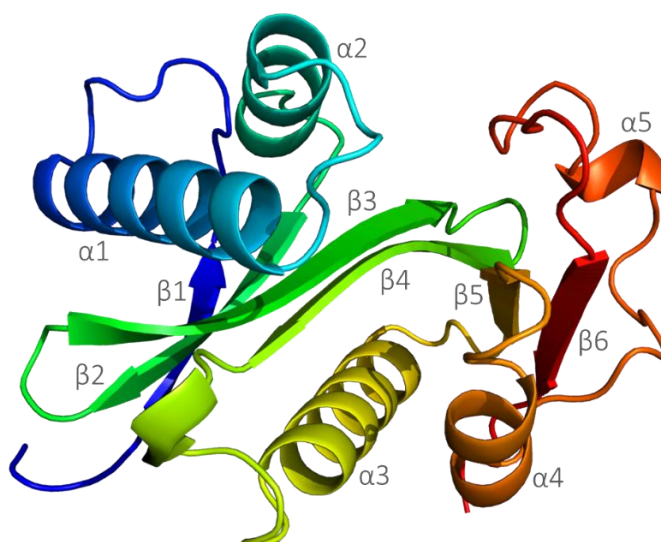
within gene families.<sup>219</sup> One such tree for the HAT domains is shown in Figure 5.1.<sup>380</sup> Common aliases of the HAT proteins are listed in Table 5.1. As shown, the domains can be separated into families. The MYST family (after yeast members MOZ, YBF2/SAS3, SAS2 and TIP60) is the largest. The p300/CBP family comprises of p300, CREBBP and ATAT1 ( $\alpha$ -tubulin N-acetyltransferase 1), and KAT2A and KAT2B are in the GNAT (GCN5-related N-acetyltransferases) family. The other HATs are transcriptional co-activators or steroid receptor co-activators, which harbour acetyltransferase activity alongside their other functions.<sup>474</sup> Interestingly, except from at the core region of the acetyl-CoA binding sites, there is little overall sequence conservation between these HAT families. This diversity should enable development of selective tool inhibitors. Within families conservation is higher. As discussed, KAT2A and KAT2B have 70% sequence identity.<sup>435</sup>

**Table 5.1. Alternative Names for Histone Acetyltransferases.** HATs listed with families, subtypes and aliases.<sup>474</sup>

Family	Subtype	Aliases
Cytoplasmic	KAT1	HAT1
	HAT4	NAA60
GNAT	KAT2A	GCN5
	KAT2B	PCAF
	KAT9	ELP3
p300/CREBBP	KAT3A	CREBBP, CBP
	KAT3B	p300, EP300
	ATAT1	MEC17
MYST	KAT5	TIP60
	KAT6A	MYST3, MOZ
	KAT6B	MYST4, MORF
	KAT7	MYST2, HBO1
	KAT8	MYST1, MOF
Transcriptional Co-activators	KAT4	TAF1, TAFII250, TFIID1
	KAT12	GTF3C4, TFIIC90
Steroid Receptor Co-activators	KAT13A	NCOA1, SRC1
	KAT13B	NCOA3, SRC3, AIB1, ACTR
	KAT13C	NCOA2, SRC2, P600
	KAT13D	CLOCK

The first HAT X-ray crystal structures resolved were *Saccharomyces cerevisiae* HAT1 (PDB ID: 1BOB)<sup>475</sup>, *Saccharomyces cerevisiae* KAT2A (PDB ID: 1YGH)<sup>476</sup> and *Homo sapiens* KAT2B (PDB ID: 1CM0).<sup>477</sup> The KAT2A HAT domain structure, which is well conserved between yeast and humans, has a mixed  $\alpha/\beta$  topology, comprising of five  $\alpha$ -helices and six  $\beta$ -strands with a globular fold, Figure 5.2.<sup>241</sup> Overall, the domain resembles a vice. The core, which forms the base of the vice, is composed of the three-stranded antiparallel  $\beta$ -sheet of  $\beta$ 2,  $\beta$ 3 and  $\beta$ 4, the  $\alpha$ -helix  $\alpha$ 3 and the strand-loop-helix containing  $\beta$ 5 and  $\alpha$ 4. The N- and C-termini form the two sides of the vice. At the N-terminus,  $\beta$ 1 contributes to the core  $\beta$ -sheet via hydrogen bonds with  $\beta$ 2, while  $\alpha$ 1 and

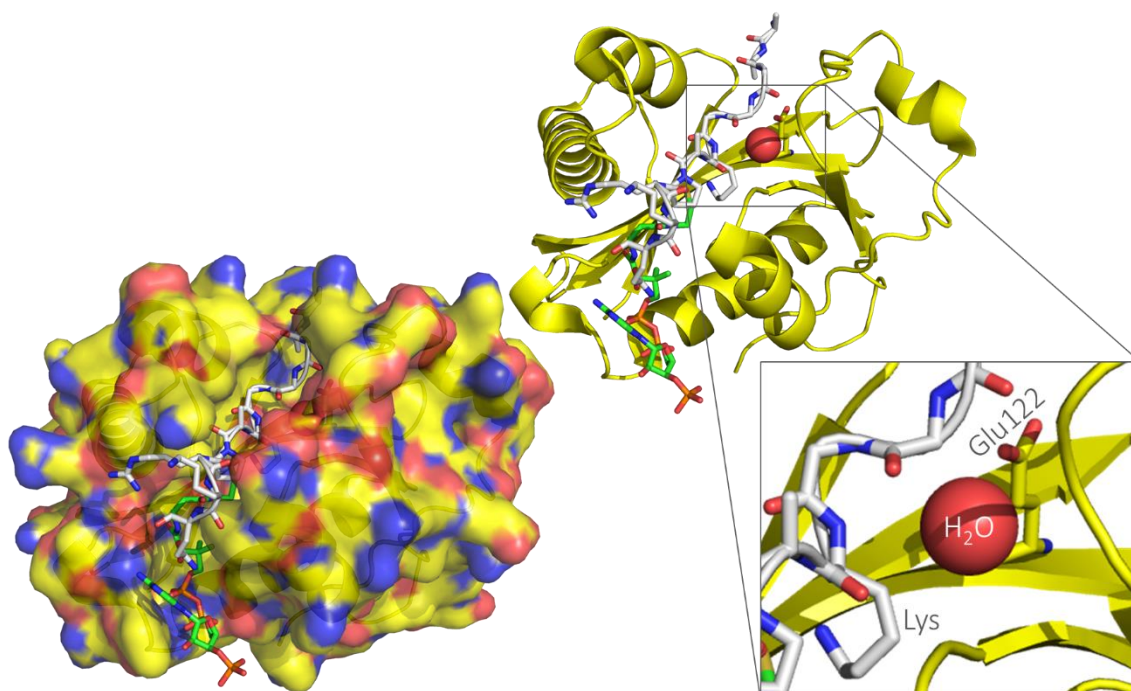
$\alpha 2$  sit off to one side and above the core. At the C-terminus, the  $\beta 6$  strand of the loop- $\alpha 5$ -loop- $\beta 6$  substructure is associated with the core by hydrogen bonding with  $\beta 5$ . The positioning of the core and the N- and C-termini creates a pronounced cleft, approximately  $10 \times 10 \times 20 \text{ \AA}$ , suitable for substrate binding. While the core domain is structurally conserved among HATs, the terminal regions show no sequence homology with other acetyltransferases.<sup>476</sup>



**Figure 5.2. KAT2A HAT Domain Topology.** Structure of *Homo sapiens* GCN5 acetyltransferase domain (PDB ID: 1Z4R).<sup>241</sup> Protein chain coloured from blue at N-terminus to red at C-terminus and  $\alpha$ -helices and  $\beta$ -strands labelled.

HATs are bi-substrate enzymes, meaning that they bind and convert two substrates during catalysis. Theoretically, there are three catalytic mechanisms that bi-substrate enzymes can employ. Firstly, a random-order ternary complex mechanism, in which both substrates bind to the enzyme, in any order, to form a ternary complex and then the acetyl is transferred directly from the acetyl-CoA to the lysine. Secondly, a compulsory-order ternary complex mechanism, which again relies on formation of a ternary complex and direct transfer of the acetyl group, but the substrates must bind in a particular order. Finally, a ping-pong mechanism, in which acetyl-CoA binds first and the acetyl is transferred to an amino acid at the enzyme catalytic site, then subsequently the lysine substrate binds and is acetylated. All three of these mechanisms require a general base at the HAT catalytic site, such as a glutamic acid residue, to facilitate nucleophilic attack at the acetyl-CoA by deprotonating the lysine residue. In addition, the ping-pong mechanism requires a residue capable of accepting the acetyl group, such as a cysteine.<sup>474</sup> There is increasing evidence that GCN5 employs a compulsory-order ternary complex mechanism. A conserved glutamic acid residue (*S. cerevisiae* Glu173/*Homo sapiens* Glu582/*Tetrahymena* Glu122) located at the bottom of the cleft in the GCN5 catalytic site is implicated as the general base necessary for catalysis.<sup>476,478</sup> Bi-substrate kinetic experiments indicated that both substrates are required to bind before catalysis, forming a ternary complex in a sequential manner, where acetyl-CoA binds first, followed by the lysine substrate.<sup>479</sup>

Also in accordance with this, Rojas *et al.* reported a crystal structure of the HAT domain of *Tetrahymena* GCN5 with coenzyme A and a histone H3 peptide bound in the cleft (PDB ID: 1QSN).<sup>480</sup> They showed that histone H3 binding is dependent on structural contributions from CoA, which reorients GCN5 upon binding. In addition, they propose that the conserved general base, Glu122, which is responsible for extracting the proton from the substrate lysine, acts via a mediating water that shuttles the proton from the lysine to the glutamic acid, Figure 5.3.<sup>480</sup>



**Figure 5.3. KAT2A HAT Domain Substrate Binding.** *Tetrahymena* GCN5 acetyltransferase domain (PDB ID: 1QSN)<sup>480</sup> illustrated with and without protein surface. KAT2A shown in yellow, CoA in green and histone H3 peptide in white, with heteroatoms highlighted. Conserved general base Glu122, lysine residue and mediating water depicted.

The diverse cellular and physiological implications of lysine acetylation do not result solely from the effect at histones. The number of non-histone proteins known to be subject to acetylation is growing rapidly. The first non-histone target discovered was tubulin<sup>481</sup>, followed some 10 years later by p53<sup>482</sup>, HIV-1 transcriptional regulator Tat<sup>483,484</sup> and nuclear factor- $\kappa$ B (NF- $\kappa$ B).<sup>485</sup> In 2006, Kim *et al.* conducted the first proteomic survey of protein acetylation. They used a screen combining immunoaffinity purification of the lysine-acetylated peptides with peptide identification by nano-HPLC/mass spectrometric analysis, to identify 388 acetylation sites in 195 proteins.<sup>486</sup> In 2009, Choudhary *et al.* elevated this approach, incorporating immunoaffinity purification, isoelectric focusing and high-resolution mass spectrometry. They identified 3600 lysine-acetylation sites in 1750 proteins.<sup>487</sup> With such an extensive list of acetylated proteins involved in an array of cellular processes, acetylation has finally been established as a globally important post-translational modification.<sup>467</sup> This abundance of interacting species and associated cellular processes might increase the likelihood that therapeutics targeting KAT2A

would suffer from adverse effects. However, it is likely that these risks would be outweighed by the benefits and therefore tolerable, given that current approved cancer therapeutics leave considerable room for improvement.<sup>10</sup> Irrespective of this, tool inhibitors of the KAT2A HAT domain will be invaluable in enabling mechanistic biological investigation and disease validation.

#### 5.1.2 Existing Inhibitors

Analysis of protein acetylation and the enzymes involved in writing and erasing acetyl marks has highlighted many novel epigenetic drug targets and thereby provoked translational research. HDACs have been studied extensively, primarily regarding their roles in cancer. To date, four pan HDAC inhibitors, Vorinostat (SAHA), Romidepsin (Istodax), Belinostat (Beleodaq) and Panobinostat (Farydak), have been approved by the FDA for the treatment of cancers such as cutaneous T-cell lymphoma, peripheral T-cell lymphoma, and multiple myeloma.<sup>488</sup> In contrast, no clinical applications have been described for HATs, despite their diverse roles in disease.<sup>474</sup>

Research into the roles of HATs has hitherto depended on genetically modified mice and immortalised cell lines. These methodologies have specific limitations. Mouse models can be unachievable, as some HATs are essential and therefore knock-out is lethal<sup>489,490</sup>, and immortalised cell lines can behave quite differently to cells *in vivo*.<sup>474</sup> There has therefore been a concerted effort in drug discovery efforts to identify small molecule tool inhibitors of HAT activity, to use as research tools and to explore the potential for therapeutic intervention.<sup>491</sup>

#### *Bi-substrate inhibitors*

Various different approaches have been utilised in developing tool inhibitors. The first class of HAT inhibitors developed were bi-substrate inhibitors, inspired by the ternary complex between the histone substrate, acetyl-CoA and HAT enzyme during the acetyl transfer. The inhibitors comprise of derivatives of CoA covalently linked to lysine substrate peptides of varying lengths. In 2000, Lau *et al.* described the design, synthesis, and application of peptide-CoA conjugates as inhibitors of p300 and PCAF.<sup>492</sup> H3-CoA-20 achieved effective and selective inhibition of PCAF HAT activity, Figure 5.4. Unfortunately, due to their partial peptide structure, these inhibitors exhibit poor metabolic stability and poor cell permeability, limiting their potential applications.

#### *Natural Products*

Several natural products have been identified as HAT inhibitors, Figure 5.4. Anacardic acid, the major component of the liquid extracted from cashew nut shells, was identified as a non-competitive HAT inhibitor in 2003.<sup>493</sup> It exhibits inhibitory activity against PCAF (IC<sub>50</sub> 5 µM), but is limited by its poor permeability and poor selectivity both between HATs and with other protein families.<sup>494</sup> Curcumin is the principal curcuminoid of turmeric and was published as a

HAT inhibitor in 2004. It demonstrates some selectivity between HATs, inhibiting the enzymatic activity of p300 and CREBBP activity but not that of PCAF.<sup>495</sup> Garcinol is a polyisoprenylated benzophenone from the dried rind of kokum fruit, *Garcinia indica*. It exhibits good cell permeability and has been shown to inhibit PCAF (IC<sub>50</sub> 5  $\mu$ M), but again is not selective and is therefore highly cytotoxic.<sup>496</sup> Recently it has been demonstrated that garcinol inhibits TgGCN5b, the nuclear acetyltransferase required for replication of *Toxoplasma tachyzoite*. Treatment with garcinol resulted in reduced global lysine acetylation, particularly at histone H3, which induced aberrant expression of TgGCN5b dependent genes, and inhibited replication.<sup>497</sup> Finally, embelin, a hydroxybenzoquinone isolated from powdered *Embelia ribes* berries, was reported as a non-competitive PCAF inhibitor. It was demonstrated that embelin inhibited the HAT activity of PCAF with some selectivity over that of p300 and TIP60, by targeting the acetyl-CoA-PCAF interaction.<sup>498</sup> Despite the limitations due to lack of selectivity within the HAT family and with other target families, these compounds are useful templates in developing HAT modulators.<sup>491</sup>

#### *Natural Product Analogues*

To date, modifying such natural product lead structures, with the aim of enhancing inhibitory properties, selectivity, and cell permeability, has proved one of the most successful approaches for developing small molecule HAT inhibitors, Figure 5.4.<sup>491</sup> Using the crystal structure of *Homo sapiens* PCAF with CoA (PDB ID: 1CM0)<sup>477</sup>, Ghizzoni *et al.* proposed a binding model for PCAF inhibition by anacardic acid and thereby designed a series of derivatives as potential novel inhibitors. One compound afforded a twofold improvement in inhibitory potency for PCAF.<sup>499</sup> A curcumin-derived, cinnamoyl analogue was identified which inhibits p300, PCAF and GCN5 acetyltransferase activity *in vitro*. However, this was not selective between HATs and it exhibited poor cell permeability.<sup>500</sup> In 2010, hydrazinocurcumin CTK7A, a water-soluble curcumin derivative, was demonstrated to inhibit p300, CREBBP and PCAF.<sup>501</sup>

#### *Synthetic Small Molecules*

Methods such as high-throughput and virtual screening have highlighted another handful of synthetic small molecules as HAT inhibitors, Figure 5.4. The  $\alpha$ -methylene- $\gamma$ -butyrolactone MB-3 has been identified as a small, cell permeable KAT2A inhibitor.<sup>502</sup> MB-3 was designed and synthesised in 2004, but it remains one of the most potent inhibitors reported for the KAT2A HAT domain (IC<sub>50</sub> 100  $\mu$ M), with affinity comparable to histone H3. Unfortunately, MB-3 exhibits limited selectivity for KAT2A over other HAT domains, such as CREBBP (IC<sub>50</sub> 500  $\mu$ M)<sup>502</sup>, and therefore does not represent a useful tool compound. MB-3 is expected to inhibit enzyme activity irreversibly, as a result of the  $\alpha,\beta$ -unsaturated carbonyl moiety accessible for Michael



addition, which may also result in poor specificity against other protein families. However, Biel *et al.* report that a time dependence study of the inhibition mode suggests non-covalent binding.<sup>502</sup> A recent study confirmed that MB-3 inhibits GCN5 HAT activity in *Arabidopsis thaliana*. *Arabidopsis thaliana* seedlings were germinated in the presence of MB-3. Three hours post-treatment H3K14 acetylation was reduced and GCN5 related genes were downregulated.<sup>503</sup> MB-3 is often employed as a reference drug in testing new potential inhibitors. In one such example, quinoline derivatives were identified from a yeast phenotypic screen, which significantly reduced *S. cerevisiae* cell growth and induced histone H3 hypoacetylation. Structural optimisation led to 4-hydroxy-2-pentylquinoline-3-carboxylic acid, which reduced the HAT catalytic activity to an extent comparable to MB-3.<sup>504</sup>

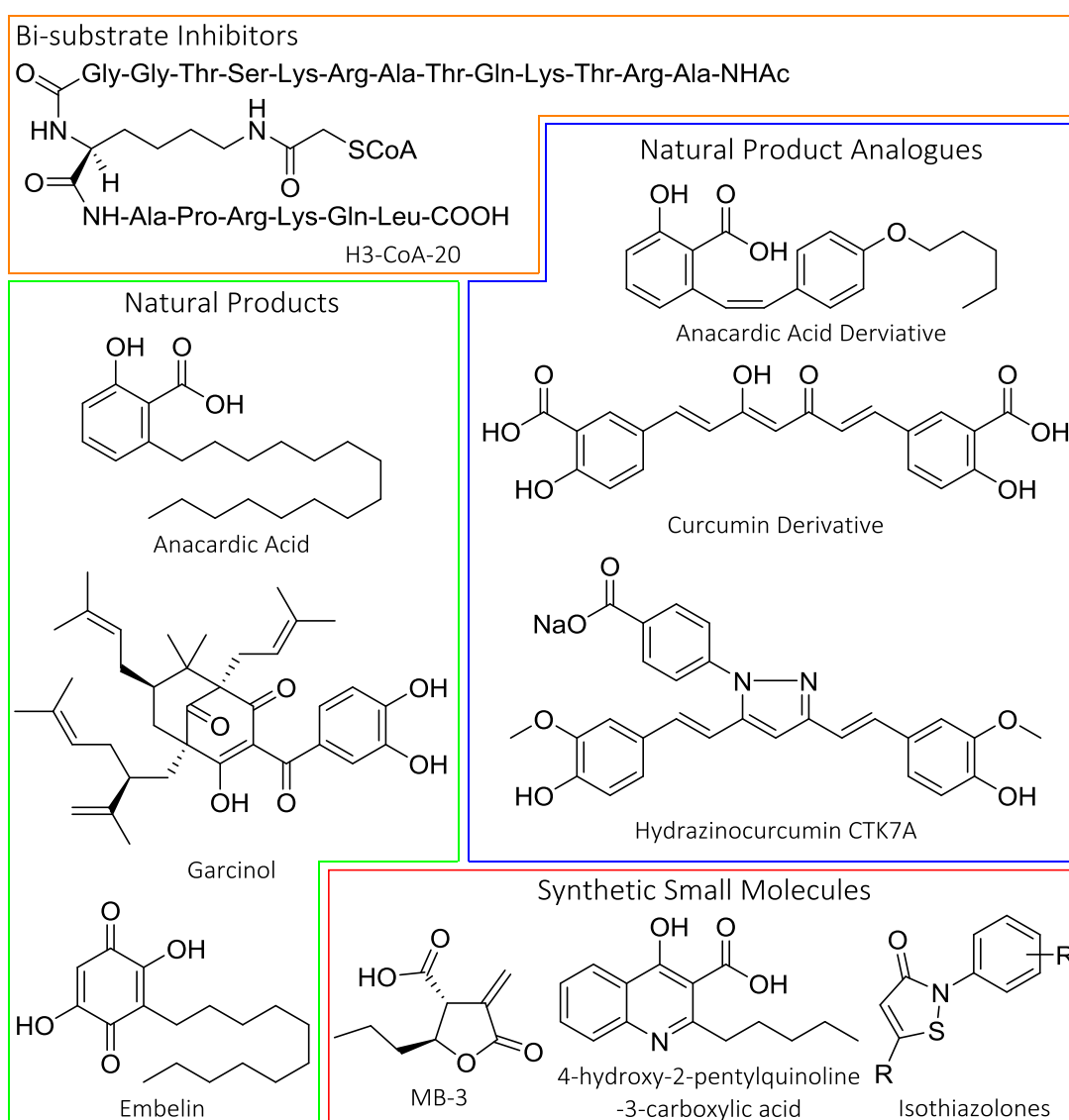


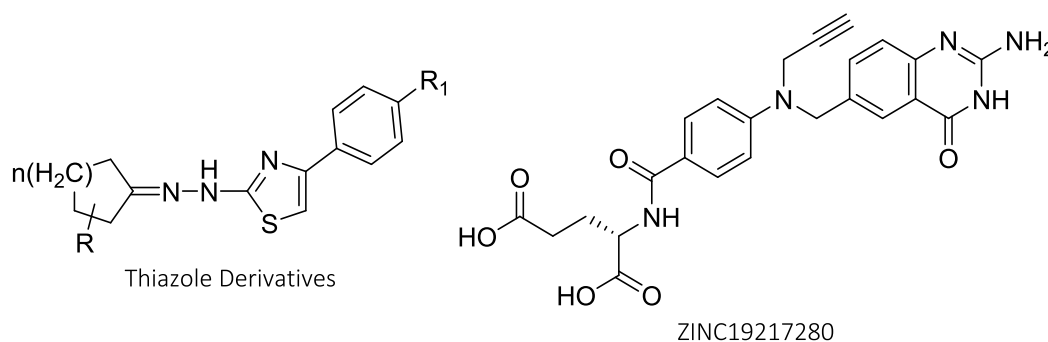
Figure 5.4. Existing KAT2A and KAT2B HAT Inhibitors.

A series of 35 isothiazolone-based p300 and PCAF inhibitors were identified using a high-throughput screen. The most promising hits were confirmed to reduce cellular acetylation in

human colon tumour cell lines and total acetylation of histones H3 and H4 and  $\alpha$ -tubulin.<sup>505</sup> It was proposed that isothiazolones irreversibly inhibit enzyme activity, binding covalently to cysteines to form disulphides via cleavage of the nitrogen-sulphur bond. The high chemical reactivity of isothiazolones is the major limitation of this class of compounds, but unfortunately this reactivity is demonstrably linked to HAT inhibitory activity.<sup>506</sup> A number of attempts have been made to identify analogues that offer reduced chemical reactivity while retaining effective HAT inhibition, and to elucidate the structure-activity relationship for isothiazolone-based inhibitors. 5-Chloroisothiazolones<sup>507</sup>, N-pyridylisothiazolones<sup>508</sup> and pyridoisothiazolones<sup>509</sup> were found to be potent PCAF inhibitors. Unfortunately though, the high reactivity of isothiazolones towards thiolates<sup>510</sup> has proved irrepressible, limiting their applicability in biological systems.<sup>474</sup>

#### *Ineffectual Inhibitors*

In 2009, Chimenti *et al.* reported a series of novel thiazole derivatives, Figure 5.5, which inhibit GCN5 HAT activity by modulating the network in which it is involved. Importantly, these compounds were selected for their ability to inhibit the growth of yeast with GCN5 deletions.<sup>511</sup> A number of studies have emanated from this research and unfortunately these thiazole derivatives have begun to be misinterpreted as GCN5 and PCAF inhibitors.<sup>512–514</sup> A recent study describes the effect of one analogue, a ‘novel PCAF and GCN5 histone acetyltransferase inhibitor’, preferentially targeting lung cancer stem-like cells derived from non-small cell lung cancer patients.<sup>515</sup> These compounds do not directly interact with KAT2A, but target the functional network via an undetermined interacting protein.



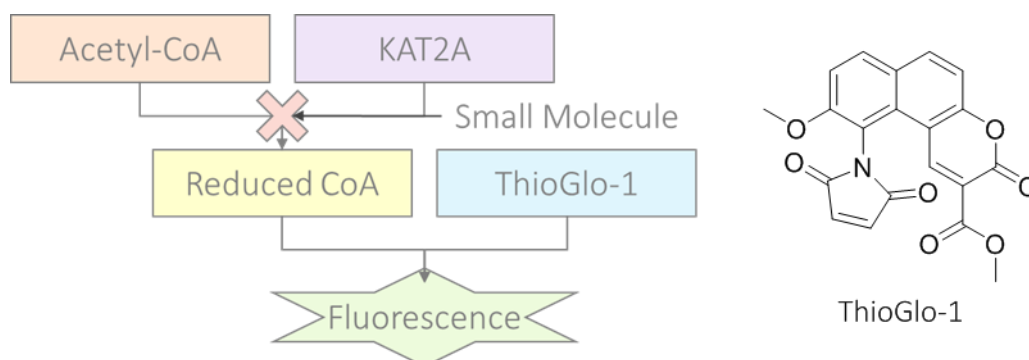
**Figure 5.5. Unsuitable HAT Inhibitors.** Thiazole derivatives modulate GCN5 network, not GCN5 directly.<sup>511</sup> ZINC19217280 selective for *Plasmodium falciparum* GCN5 and do not inhibit *Homo sapiens* homologue.<sup>516</sup>

In a recent study, 20 potential inhibitors of *Plasmodium falciparum* GCN5 were reported and it was experimentally validated that ZINC19217280 possesses antimalarial activity, Figure 5.5. While the *in vitro* parasite growth assay indicated an  $IC_{50}$  in the nanomolar range ( $\sim 225$  nM), no effect was observed in mammalian fibroblast cells. It was therefore resolved that ZINC19217280 selectively inhibits *Plasmodium falciparum* GCN5 and not the *Homo sapiens* homologue.<sup>516</sup>

Overall, the development of HAT tool inhibitors has unfortunately proved very challenging. The current inhibitors are largely non-selective<sup>494,496</sup> or have poor permeability<sup>492</sup> and many are redox cyclers or covalent modifiers<sup>506</sup>. As a result there remains a crucial need for new potent and selective HAT inhibitors with improved therapeutic potential.<sup>474</sup>

### 5.1.3 New Chemical Matter

As well as the known HAT inhibitors, a ChEMBL<sup>159</sup> search for small molecules bioactive against KAT2A yielded an array of hits from a high-throughput screen described in the PubChem BioAssay database (PubChem BioAssay AID: 504327).<sup>4</sup> 793 bioactive compounds were identified in this screen, completed by the NIH Chemical Genomics Center (NCGC). A corresponding hit confirmation assay was reported (PubChem BioAssay AID: 588347)<sup>5</sup>, again by the NCGC, which confirmed the activity of 96 of these compounds. In the activity assay, a pro-fluorescent maleimide derivative, ThioGlo-1, was used to measure the production of reduced CoA. ThioGlo-1 fluoresces upon reaction with free thiols, such as that in reduced CoA. Inhibition of HAT activity by a small molecule prevents the reduction of acetyl-CoA to the thiol, and therefore ThioGlo-1 will not fluoresce, Figure 5.6. When the screens were conducted, an initial read was carried out in attempt to identify compounds with fluorescence quenching properties, which would be likely to afford false positives. However, the authors anticipate that a large number of artefacts will remain and recommend counter-screens and further assay validation.



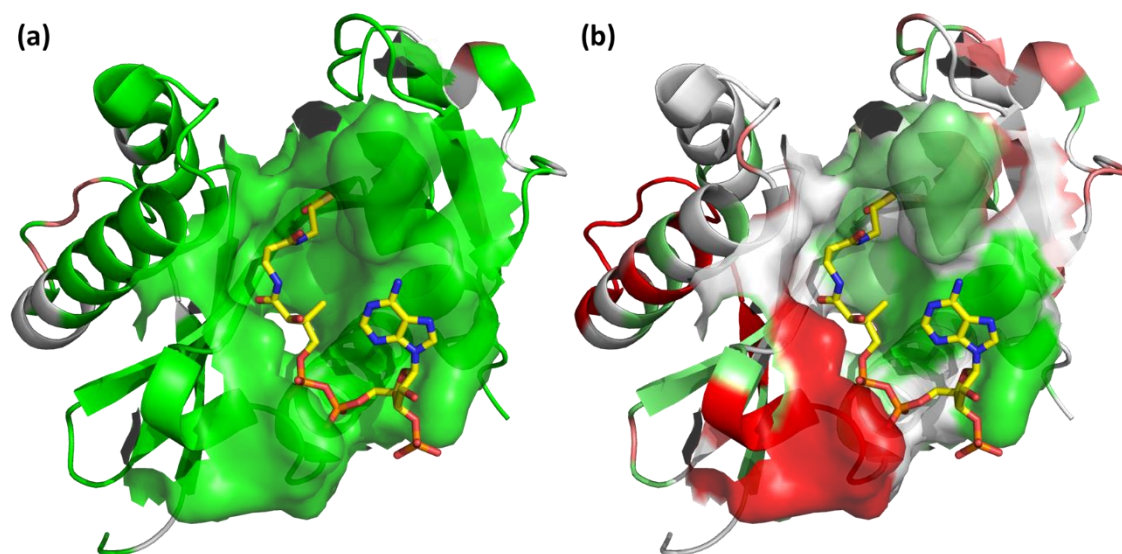
**Figure 5.6. ThioGlo-1 and Schematic of Associated Fluorescence Assay.** Assay utilised in HTS reported in the PubChem BioAssay database (PubChem BioAssay AID: 504327).<sup>4</sup>

## 5.2 Results

### 5.2.1 Selectivity Potential

In Chapter 4.2, a druggable structural pocket was identified at the KAT2A HAT domain using Fpocket 2.0<sup>165,170</sup>, which corresponds to the binding position of acetyl-CoA. EMBL-EBI Clustal Omega<sup>181,182</sup> was used to compile pairwise and multiple sequence alignments in attempt to assess the potential for selectivity at this site, Figure Apx.5. Those amino acids which comprise the KAT2A HAT domain were selected and Scorecons<sup>220</sup> was used to rate the conservation at

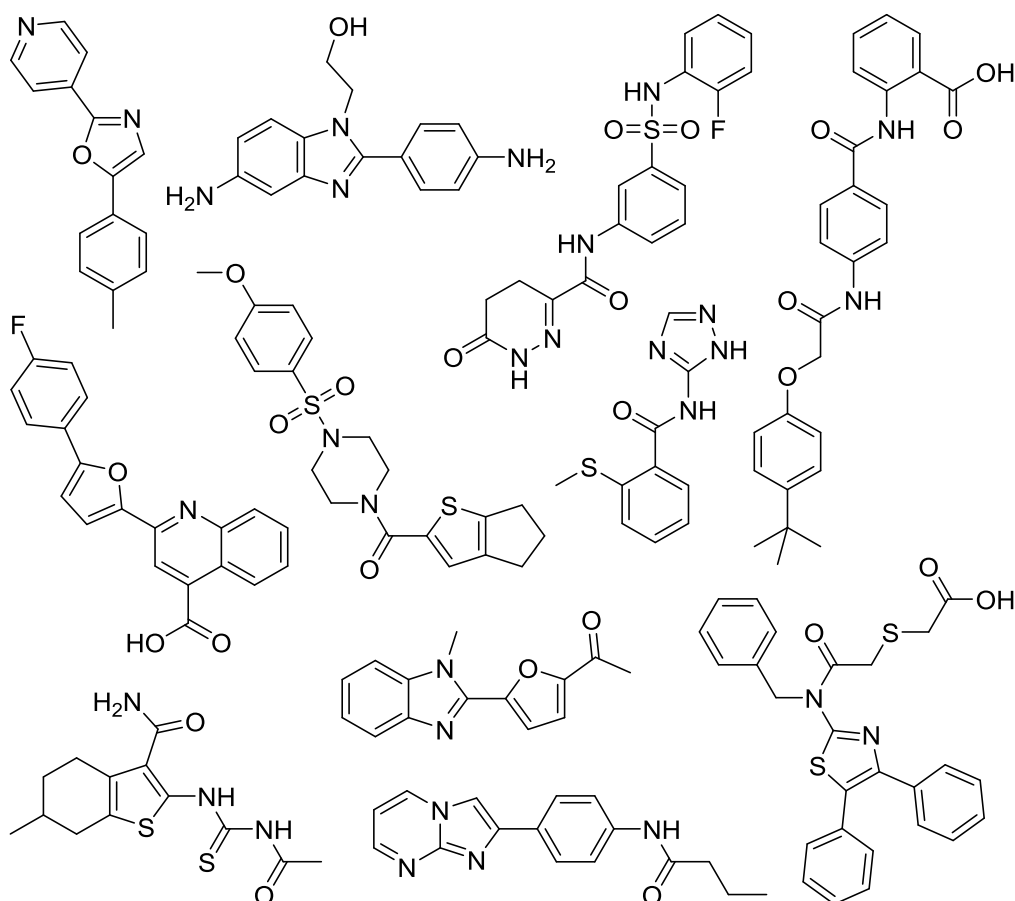
each associated site. Predictably, KAT2A and KAT2B exhibit extremely high homology, Figure 5.7 (a), but fortuitously the residues are poorly conserved across other HAT families, Figure 5.7 (b). There is therefore a high chance of achieving tool inhibitors that are selective for the two GCN5 homologues, but not that can distinguish between them. This was not discouraging. Small molecule dual inhibitors will nevertheless represent useful tools, which will enable biological exploration of the target proteins to further elucidate their biological roles and help in establishing functional assays.



**Figure 5.7. Assessing Residue Conservation at KAT2A Acetyl-CoA Binding Site.** *Homo sapiens* GCN5 Acetyltransferase domain (PDB ID: 1Z4R)<sup>241</sup> with acetyl-CoA in yellow and KAT2A coloured to indicate amino acid conservation determined by Scorecons<sup>220</sup>; red indicates the least conserved regions and bright green the most. (a) Results from pairwise alignment of KAT2A and KAT2B. (b) Results from multiple sequence alignment of GNAT and MYST families.

### 5.2.2 Sifting the Existing Chemical Matter

As discussed, a ChEMBL<sup>159</sup> search for small molecules bioactive against KAT2A yielded (among other entries) a high-throughput screen, described in the PubChem BioAssay database (PubChem BioAssay AID: 504327)<sup>4</sup> and corresponding hit confirmation assay (PubChem BioAssay AID: 588347)<sup>5</sup> completed by the NCGC. KNIME<sup>242</sup> was utilised in filtering the results of the confirmatory screen. KNIME, the Konstanz Information Miner, is a modular environment, which enables visual assembly and interactive execution of a data pipeline.<sup>242</sup> A KNIME workflow was designed to filter the results and identify compounds that were defined as ‘active’, had acceptable curve descriptions, discarding any with efficacy >150 calculated from a partial curve, which did not reach the second asymptote, and were available from preferred suppliers, such that they could be purchased easily and at low cost. 47 compounds remained after filtering. Attempts to cluster these into 5-10 series of 3-5 compounds was unsuccessful due to the small size of the set. Instead, 11 compounds were manually selected that were drug-like and avoided undesirable motifs, such as pan-assay interference compounds (PAINS)<sup>517</sup>, Figure 5.8.

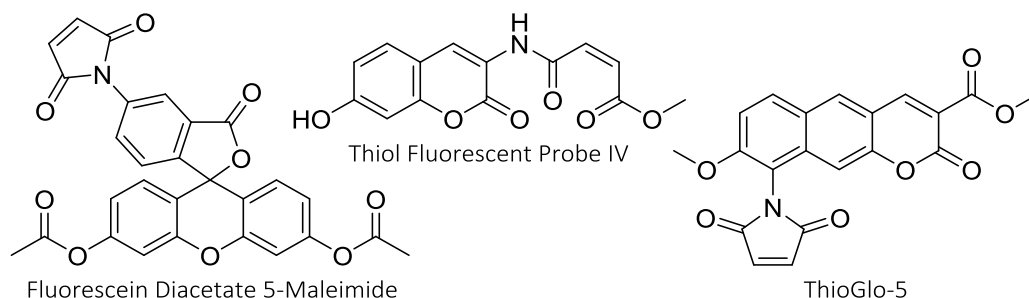


**Figure 5.8. Prospective KAT2A HAT Inhibitors.** From NCGC hit confirmation assay (PubChem BioAssay AID: 588347).<sup>5</sup>

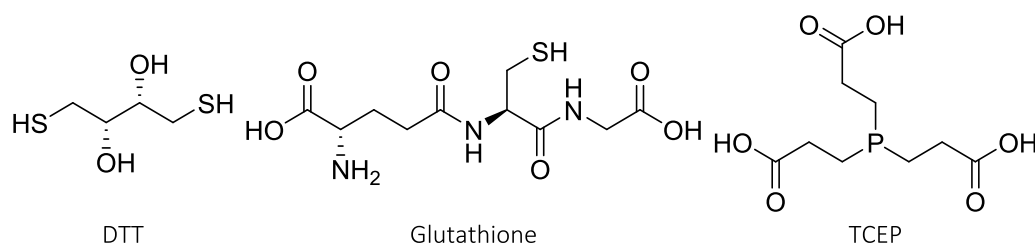
### 5.2.3 Fluorescence Based Activity Assay

Having established that the KAT2A HAT domain represents a desirable target for specific tool inhibitor development and identified the HTS (PubChem BioAssay AID: 504327)<sup>4</sup> reported by the NCGC, attempts to establish this assay in house and confirm these results were prioritised.

Unfortunately, since the HTS was reported, ThioGlo-1 had been discontinued. Three alternative thiol reactive fluorescent probes were trialled in this work, fluorescein diacetate 5-maleimide, Thiol Fluorescent Probe IV and ThioGlo-5, Figure 5.9.

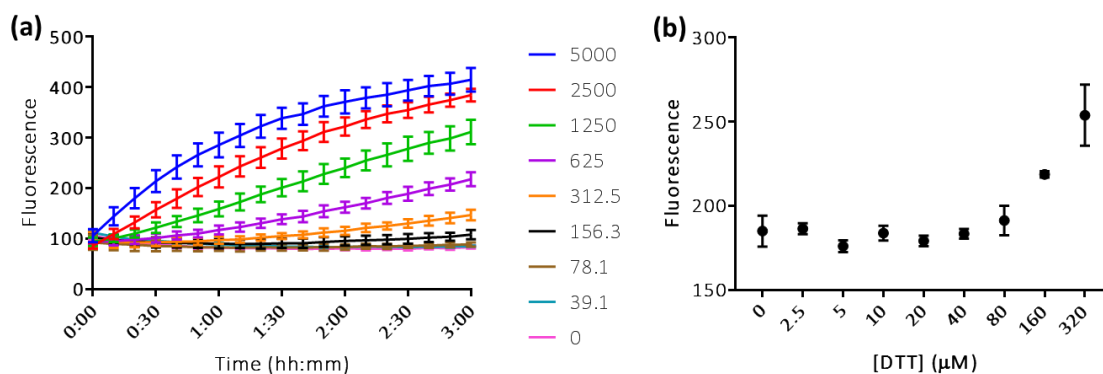


**Figure 5.9. Thiol Reactive Fluorescent Probes.** Fluorescein diacetate 5-maleimide, Thiol Fluorescent Probe IV and ThioGlo-5.



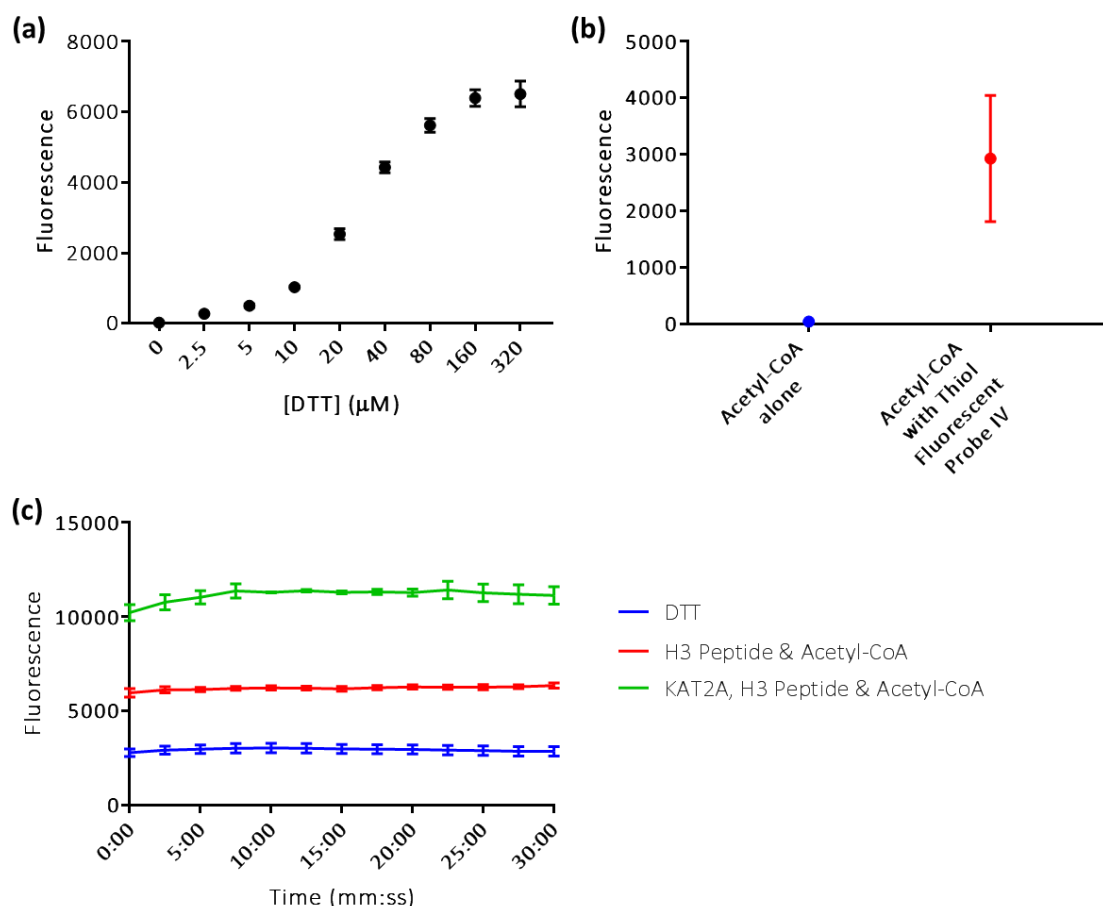
**Figure 5.10. Common Reducing Agents.** Reducing agents used in assay buffers to stabilise the protein and prevent formation of undesirable disulphide bonds.

To begin with, the sensitivity of fluorescein diacetate 5-maleimide was assessed. 50  $\mu\text{M}$  fluorescent probe was added to solutions of dithiothreitol (DTT), Figure 5.10, at a range of concentrations, and the developing fluorescence was measured over time, Figure 5.11 (a). It was determined that the fluorescein diacetate 5-maleimide was slow acting, which was not ideal but could be tolerated, requiring a delay between indicator addition and plate reading, but also that it may not be sufficiently sensitive to measure the necessary thiol concentrations. The experiment was repeated with increased fluorescent probe and decreased DTT, to mimic the amount of reduced CoA expected in the final assay conditions, Figure 5.11 (b). The maximum expected concentration of reduced CoA was calculated at 70  $\mu\text{M}$ , assuming that the protocol used in the HTS (PubChem BioAssay AID: 504327)<sup>4</sup> would be mimicked and that the acetyl-CoA would be completely converted to reduced CoA. There was no assay window at these thiol concentrations, confirming that the fluorescein diacetate 5-maleimide was not sufficiently sensitive to use in this assay.



**Figure 5.11. Sensitivity of Fluorescein Diacetate 5-Maleimide.** (a) Fluorescence of 50  $\mu\text{M}$  probe with 0-5000  $\mu\text{M}$  DTT, 40  $\mu\text{L}$  well volume, measured for three hours. (b) Fluorescence of 127  $\mu\text{M}$  probe with 0-320  $\mu\text{M}$  DTT, 40  $\mu\text{L}$  well volume, measured after three hour incubation. Graphs represent mean  $\pm$  1 SD of three experiments.

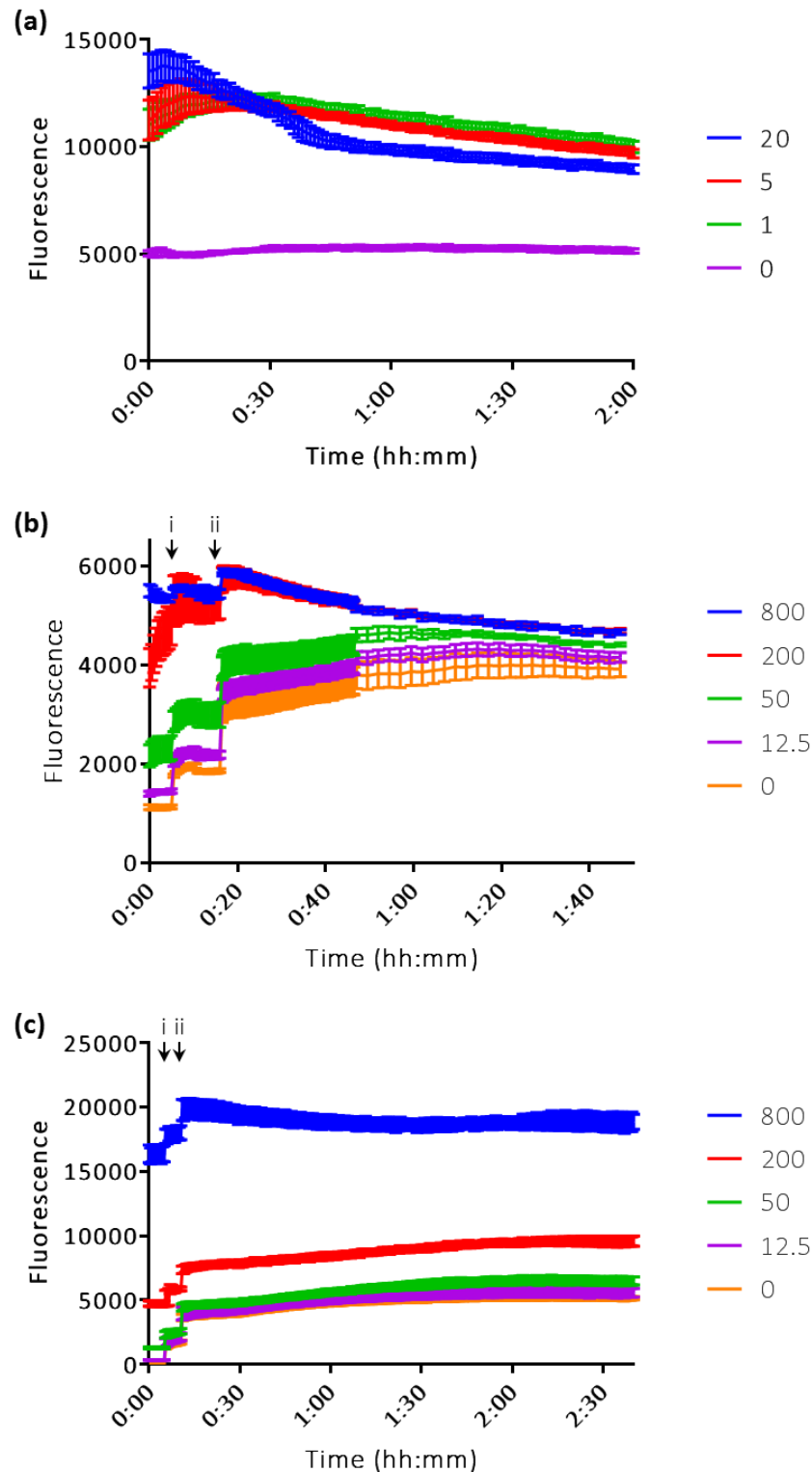
With fluorescein diacetate 5-maleimide proving unsuitable, attention turned to Thiol Fluorescent Probe IV. Favourably, this was much faster acting, with the reaction complete as soon as the plate was read, simplifying and accelerating experiments. The sensitivity of Thiol Fluorescent Probe IV was assessed at varying concentrations of the indicator and of DTT. This probe exhibited much improved sensitivity at concentrations as low as 10  $\mu\text{M}$ , Figure 5.12 (a).



**Figure 5.12. Preliminary Tests with Thiol Fluorescent Probe IV.** (a) Fluorescence of 10  $\mu\text{M}$  probe with 0–320  $\mu\text{M}$  DTT, 40  $\mu\text{L}$  well volume. (b) Fluorescence of 70  $\mu\text{M}$  acetyl-CoA with or without 10  $\mu\text{M}$  probe, 40  $\mu\text{L}$  well volume. (c) Fluorescence of 10  $\mu\text{M}$  probe with 70  $\mu\text{M}$  DTT or 70  $\mu\text{M}$  acetyl-CoA and 75  $\mu\text{M}$  histone H3 peptide (1–21) with or without 20 nM KAT2A, 40  $\mu\text{L}$  well volume, measured for 30 minutes. Graphs represent mean  $\pm$  1 SD of three experiments.

In another preliminary test, Thiol Fluorescent Probe IV was added to a mix of acetyl-CoA and histone H3 peptide (1–21), with and without KAT2A, and the developing fluorescence was measured for 30 minutes, Figure 5.12 (c). Compared with the 70  $\mu\text{M}$  DTT controls, it appeared that acetyl-CoA and the H3 peptide resulted in significant fluorescence. It was confirmed that acetyl-CoA induces fluorescence, Figure 5.12 (b). This could suggest either that a percentage of free thiol is permanently present in the acetyl-CoA solution, or that the fluorescent probe also interacts with unmodified acetyl-CoA. There was also an observed increase in fluorescence upon KAT2A addition, which might be attributable to protein activity. However, the KAT2A storage buffer incorporated both DTT and glutathione, Figure 5.10, and the protein contains cysteine residues, all of which contain thiols that could interact with the probe to induce fluorescence.

An attempt to measure enzyme progression at varying KAT2A concentrations over the course of two hours showed an unexpected decrease in fluorescence over time, Figure 5.13 (a). However, a slight increase was observed over the first 5–10 minutes, which could theoretically be indicative of activity, although it is more likely that the solutions were proceeding to equilibrium.



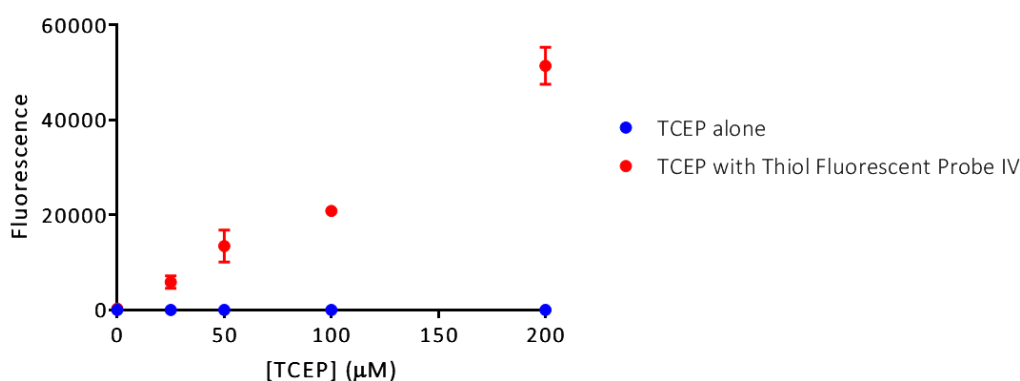
**Figure 5.13. KAT2A Activity Tests.** (a) Fluorescence of 10  $\mu$ M Thiol Fluorescent Probe IV, 70  $\mu$ M acetyl-CoA, 75  $\mu$ M histone H3 peptide (1-21) and 0-20 nM KAT2A, 40  $\mu$ L well volume, measured for two hours. (b) Fluorescence of 10  $\mu$ M probe with 0-800 pM KAT2A with stepwise addition of: (i) 70  $\mu$ M acetyl-CoA at 5 minutes and (ii) 75  $\mu$ M histone H3 peptide (1-21) at 15 minutes; 40  $\mu$ L well volume, measured for two hours. (c) Fluorescence of 40  $\mu$ M probe with 0-800 pM KAT2A with stepwise addition of: (i) 75  $\mu$ M histone H3 peptide (1-21) at 5 minutes and (ii) 70  $\mu$ M acetyl-CoA at 10 minutes; 40  $\mu$ L well volume, measured for 2.5 hours. Graphs represent mean  $\pm$  1 SD of three experiments.

To improve understanding, the activity test was repeated at lower KAT2A concentrations with stepwise addition of each assay component, Figure 5.13 (b). Initially the Thiol Fluorescent Probe



IV was added to KAT2A alone. The fluorescence was higher at increased protein concentrations, likely due to the presence of cysteine residues, or DTT and glutathione in the storage buffer. Addition of acetyl-CoA at 5 minutes resulted in a sharp increase in fluorescence at all enzyme concentrations, as expected from the previous result. Introduction of the histone H3 peptide (1-21) at 15 minutes also resulted in a similar increase in fluorescence. This was surprising, as the peptide solution contained no reducing agents. However, it was discovered that whilst the histone H3 peptide (1-21) sequence contains no cysteine residues, the commercial peptide incorporates a tag containing a cysteine. This cysteine can be expected to interact with the thiol probe. Importantly, in the 10 minutes directly following peptide addition, there was no time dependent increase in fluorescence, implying an absence of enzyme activity.

In light of the unexpectedly high thiol concentrations, it was proposed that the thiol probe was limiting the reaction. A second activity test was conducted at higher probe concentration, again with stepwise addition of the assay components, Figure 5.13 (c). This appeared to avert the decrease in fluorescence with time. Conceivably, if thiols are present in large excess, Thiol Fluorescent Probe IV could react more than once. If this final product is not fluorescent at the same wavelength as the product of the single 1,4-addition to the Michael acceptor, this could result in transitory fluorescence. However, still no enzyme activity was observed.

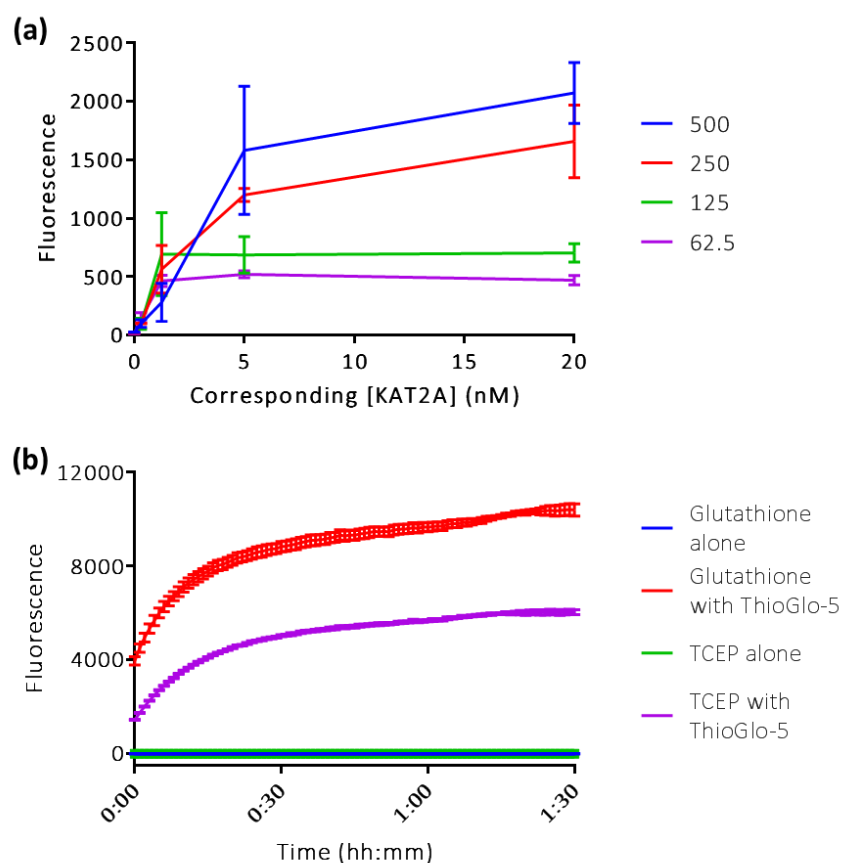


**Figure 5.14. Thiol Fluorescent Probe IV with TCEP.** Fluorescence of 0-200 μM TCEP with or without 500 μM Thiol Fluorescent Probe IV, 40 μL well volume. Graph represents mean  $\pm$  1 SD of three experiments.

Whilst the probe had likely been rate limiting at the higher KAT2A concentrations, this did not explain the lack of observable protein activity throughout. Considering that DTT and glutathione are present in the buffer to stabilise the protein and prevent formation of undesirable disulphide bonds, it was proposed that if these reducing agents react with the probe, this could leave KAT2A vulnerable to denaturation. Tris(2-carboxyethyl)phosphine (TCEP), Figure 5.10, can be utilised as an alternative reducing agent. TCEP does not contain a thiol and therefore should not react with the fluorescence indicator and could be substituted for the DTT and glutathione, to reduce background fluorescence and maintain the stability of the protein. However, unfortunately it

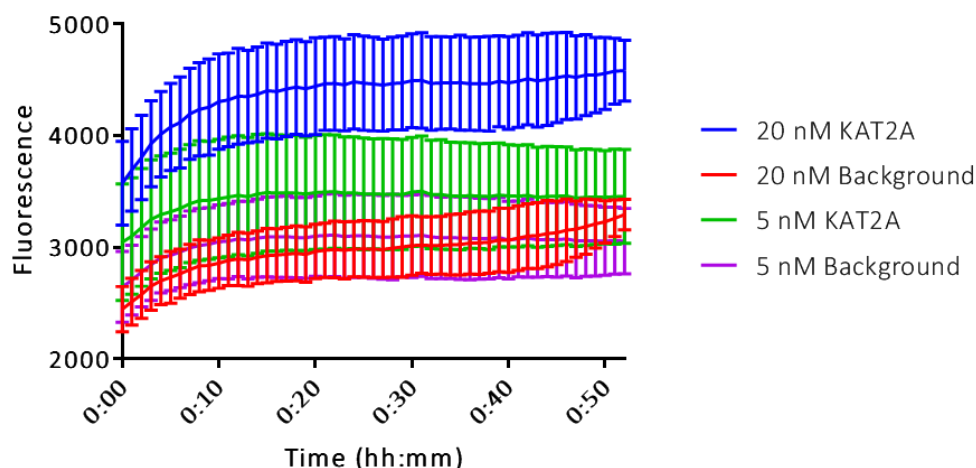
was discovered that TCEP also interacts with Thiol Fluorescent Probe IV, Figure 5.14. Thiol Fluorescent Probe IV was therefore determined unsuitable for use in the assay.

After this lack of success with Thiol Fluorescent Probe IV, attention turned to a third fluorescence indicator, ThioGlo-5. To ensure that sufficient ThioGlo-5 was used so as not to be limiting, the capacity of the probe was tested. ThioGlo-5 was added to solutions of DTT and glutathione intended to mimic the background thiol concentrations associated with 0-20 nM KAT2A in buffer, Figure 5.15 (a). Above 250  $\mu$ M ThioGlo-5 the probe did not significantly limit the fluorescence. A second preparatory experiment confirmed that ThioGlo-5 is slower acting than Thiol Fluorescent Probe IV, reacting gradually over 1.5 hours, and that like Thiol Fluorescent Probe IV, it fluoresces in the presence of TCEP, Figure 5.15 (b).



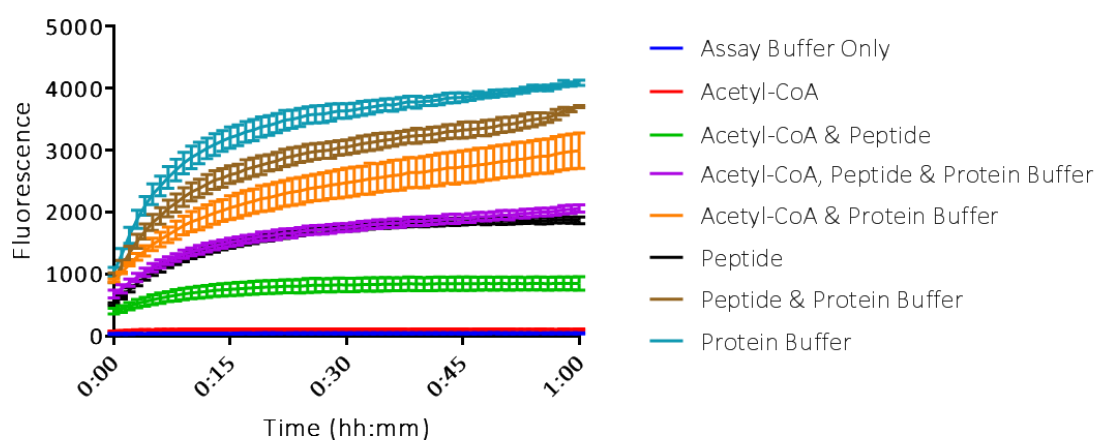
**Figure 5.15. Preliminary Tests with ThioGlo-5.** (a) Fluorescence of 62.5-500  $\mu$ M ThioGlo-5 with DTT and glutathione to mimic background thiol concentrations in 0-20 nM KAT2A, 40  $\mu$ L well volume, measured after 20 minute incubation. (b) Fluorescence of 250  $\mu$ M ThioGlo-5, with or without 192  $\mu$ M glutathione or 192  $\mu$ M TCEP, 40  $\mu$ L well volume, measured over 1.5 hours. Graphs represent mean  $\pm 1$  SD of three experiments.

An attempt to measure enzyme progression at 5 nM and 20 nM KAT2A, showed an increase in fluorescence upon protein addition, Figure 5.16. However, there was substantial error in the measurements and it was judged that the increase could plausibly be attributed to KAT2A cysteine residues. It was therefore concluded that there was no indication of enzyme activity.



**Figure 5.16. KAT2A Activity Test.** 250  $\mu$ M ThioGlo-5, 70  $\mu$ M acetyl-CoA and 75  $\mu$ M histone H3 peptide (1-21) with 0-20 nM KAT2A, employing DTT and glutathione as necessary to imitate protein storage buffer, 40  $\mu$ L well volume, measured over one hour. Graph represents mean  $\pm$  1 SD of three experiments.

A troubleshooting experiment was conducted in which every assay component was tested with ThioGlo-5, individually and in combination with every other, Figure 5.17, in the hope of identifying the most problematic elements. As expected, the DTT- and glutathione-containing protein buffer and the H3 peptide demonstrated significant fluorescence. However, introducing additional acetyl-CoA to either of these or combining the two resulted in a decrease in overall fluorescence. This was highly surprising, as it had been anticipated that any increase in thiol concentration would provoke an increase in fluorescence. This may result from ThioGlo-5 becoming limiting, or the various thiols might interact with one another. Irrespective of the explanation, it was clear that the assay was unreliable and there were many complex component interactions and factors that would impede assay optimisation. As such, the fluorescence based activity assay was abandoned.

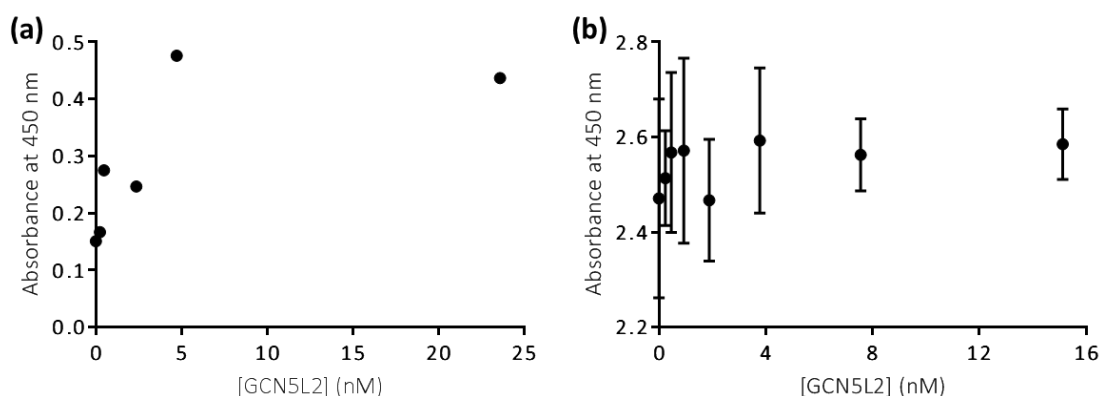


**Figure 5.17. Troubleshooting Experiment.** Fluorescence of 250  $\mu$ M ThioGlo-5 with all assay components: 70  $\mu$ M acetyl-CoA, 75  $\mu$ M histone H3 peptide (1-21), and protein storage buffer; in every combination, 40  $\mu$ L well volume, measured over one hour. Graph represents mean  $\pm$  1 SD of three experiments.

Shortly after reaching this decision, the authors of the HTS shared with us that despite their initial publication they were also unable to establish a robust assay and they planned to investigate alternative radioligand binding techniques.

#### 5.2.4 EpiQuik™ HAT Activity Assay

Having abandoned attempts to establish the fluorescence based HAT activity assay in house, and with a desire to avoid other assays dependent on production of reduced CoA, the EpiQuik™ HAT activity assay (Epigentek, UK) was identified. In this assay, the histone substrate is first stably captured in the wells. GCN5L2 and acetyl-CoA are then added to acetylate the substrate. After incubation GCN5L2 and acetyl-CoA are washed away, preventing further acetylation. A high affinity anti-acetylated histone antibody is added, the ‘capture’ antibody, which specifically recognises and binds the acetylated substrate. Any excess after incubating is washed away. A second, ‘detection’ antibody is then added, and again the excess is washed away after incubating. Finally, developing and stop solutions are applied sequentially, to afford absorbance at 450 nm. In this way, the amount of acetylated histone, which is directly proportional to the enzyme activity, is colourimetrically quantified via an enzyme-linked immunosorbent assay (ELISA)-like reaction.



**Figure 5.18. GCN5L2 Activity Measured by EpiQuik™ HAT Assay.** HAT activity proportional to absorbance measured at 450 nm using commercially available colourimetric assay kit (Epigentek, UK). (a) Absorbance measured at 0–24 nM GCN5L2. Graph represents single experiment. (b) Absorbance measured at 0–16 nM GCN5L2. Graph represents mean  $\pm$  1 SD of three experiments.

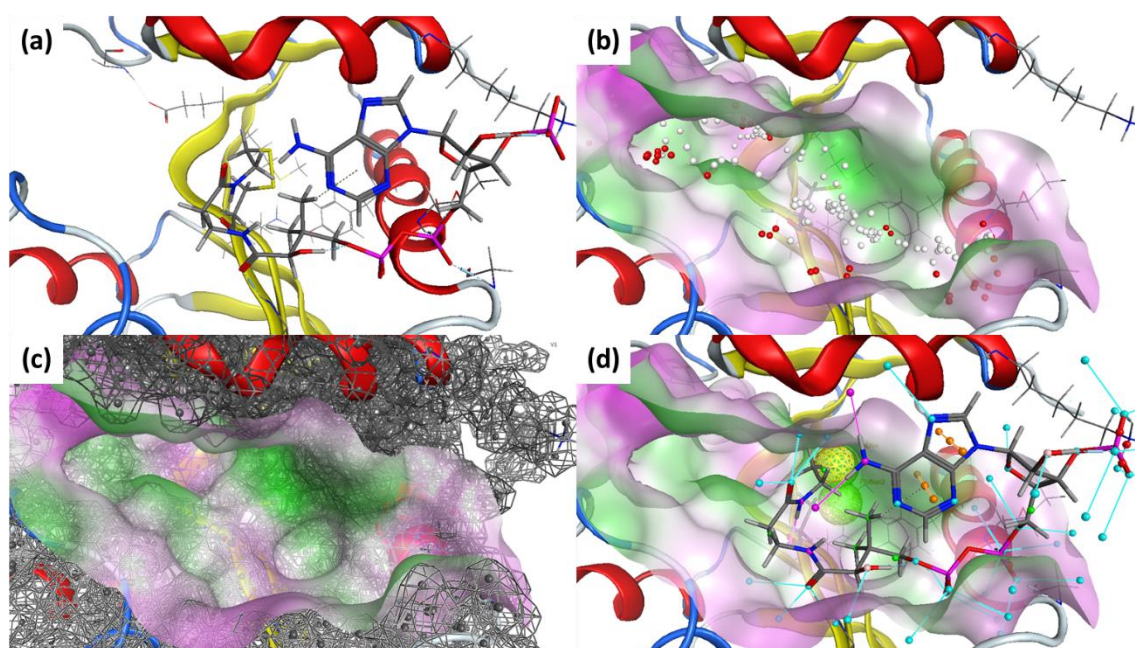
The first trial of the assay looked promising, with absorbance increasing and finally plateauing with increased GCN5L2, indicating that acetylation increases with GCN5L2 concentration until a saturation point at which the peptide is fully converted within the period of incubation, Figure 5.18 (a). However, attempts to repeat this experiment in triplicate highlighted a high margin of error, such that the standard deviation was often significantly larger than the assay window, Figure 5.18 (b). It is anticipated that by using a higher concentration of the histone substrate,

and thereby increasing the final absorbance, the assay window could be expanded, although care should be taken not to exceed the binding limit of the strip wells.

### 5.2.5 Isothiazolones as Potential HAT Inhibitors

Meanwhile, having determined that the hits from the HTS were unlikely to represent true KAT2A inhibitors, it was necessary to identify other relevant published chemical matter. The list of small molecules reportedly active against KAT2A had been exhausted, but given the extremely high homology of KAT2A and KAT2B, inhibitors of KAT2B should also exhibit activity against KAT2A. The majority of relevant compounds identified in a ChEMBL<sup>159</sup> search for KAT2B inhibitors were isothiazolones.<sup>505–509</sup>

Compounds can be docked into protein crystal structures in attempt to ascertain their binding pose and improve understanding of their activity. The *Homo sapiens* KAT2A HAT domain (PDB ID: 1Z4R)<sup>241</sup> was prepared for docking in MOE.<sup>187</sup> Firstly, it was protonated, corrected and minimised, Figure 5.19 (a). Then the Site Finder functionality was used to identify the site available for ligand binding, dummies were created at the alpha sphere centres, and a molecular surface was generated within 4.5 Å of these dummies, Figure 5.19 (b). Finally, a pharmacophore was generated to exclude a volume of radius 1 around all receptor atoms, Figure 5.19 (c).



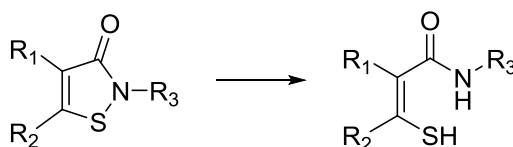
**Figure 5.19. Preparatory Work for Docking at KAT2A HAT Domain.** *Homo sapiens* KAT2A HAT domain acetyl-CoA binding site (PDB ID: 1Z4R).<sup>241</sup> (a) Prepared structure in complex with acetyl-CoA. (b) Ligand binding site depicted by dummy atoms with associated surface coloured by lipophilicity. (c) Pharmacophore excluding volume around all receptor atoms. (d) Acetyl-CoA inspired pharmacophore with thiol donor and donor projection towards cysteine. Images produced in MOE.<sup>187</sup>

The library of isothiazolones reported as KAT2B inhibitors were loaded into MOE<sup>187</sup>, prepared, protonated and minimised. The compounds were then docked into the prepared structure of

the KAT2A HAT domain using the induced fit protocol and employing the pharmacophore as a guide. Unfortunately, this conventional approach of docking the inhibitors into the structure of the KAT2A HAT domain was unsuccessful.

As discussed, the isothiazolones inhibit KAT2B irreversibly via covalent thiol interactions<sup>505</sup> with cysteine residues.<sup>509</sup> Therefore, in an attempt to improve the results, a KNIME<sup>242</sup> workflow was designed to use before docking, employing a JChem reactor node, JChem 6.1.0, 2013, ChemAxon (<http://www.chemaxon.com>), to convert the isothiazolones to 3-mercaptoacrylamides, as they will eventually reside at the active site, Figure 5.20. Additionally, the pharmacophore was elaborated to also include the acetyl-CoA thiol donor with radius 1.2 and a donor projection with radius 1.2 towards the cysteine residue at the ligand binding site, Figure 5.19 (d).

The library of mercaptoacrylamides were loaded into MOE<sup>187</sup>, prepared, protonated and minimised and then docked into the prepared structure of the KAT2A HAT domain using the induced fit protocol and employing the improved pharmacophore as a guide. A greater number of the compounds were docked into the structure using this method but there remained a lack of confidence in the results.



**Figure 5.20. Converting Isothiazolones to Mercaptoacrylamides.** Compounds converted using reactor node, JChem 6.1.0, 2013, ChemAxon ([www.chemaxon.com](http://www.chemaxon.com)) in KNIME.<sup>242</sup>

With perseverance and guidance from the CCG (Chemical Computing Group), this challenge could be surmounted and the series of isothiazolones could be docked into KAT2A. However, the isothiazolones are well known PAINS.<sup>517</sup> They are frequently identified as artefactual hits because they are highly reactive, which causes false positive signals across a variety of assays. In this particular instance the isothiazolones might represent true KAT2B inhibitors, which inhibit activity via a specific covalent interaction with the protein. However, this is unlikely and was not deemed to be worth the risk. Furthermore, the authors responsible for first identifying the isothiazolones as KAT2B inhibitors<sup>505,508</sup> confirmed that they had also found these compounds to be highly reactive and non-specific. As such, it was decided that the isothiazolones were not worth pursuing. This left no suitable chemical matter available for the KAT2A/B HAT domains.

### 5.3 Discussion

As discussed in Chapter 4.1, KAT2A is synthetic lethal with BAF180<sup>3</sup>, which is frequently mutated in clear cell renal cell carcinoma.<sup>376</sup> This protein therefore represents an interesting target for tool inhibitor development, to enable biological investigation and potentially initiate drug

discovery projects towards treating ccRCC. KAT2A contains a HAT domain and a bromodomain. In Chapter 4.2, druggable pockets were identified in both of these structural domains. As the catalytically active domain, the HAT domain was prioritised as the preferred target.

HATs represent exciting targets in drug discovery, with many potential disease applications, and are therefore interesting targets for tool inhibitor development, especially as existing HAT inhibitors are largely non-selective<sup>494,496</sup>, have poor permeability<sup>492</sup> or are covalent modifiers<sup>506</sup>, and are therefore inadequate for biological investigation.

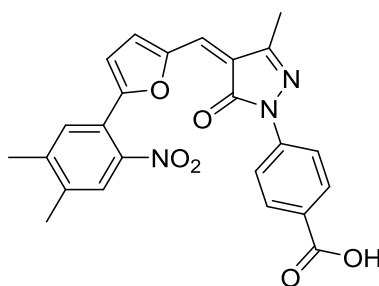
Initially, sequence alignments were used to assess the potential for selectivity. It was found that KAT2A shows extremely high homology with its closest relative, KAT2B, but that overall the residue conservation across the HAT families is poor. This should afford good selectivity in general, but it is anticipated that any small molecule inhibitor will act at both KAT2A and KAT2B.

Attempts to establish a fluorescence based activity assay, previously utilised in HTS for KAT2A HAT inhibitors (PubChem BioAssay AID: 504327)<sup>4</sup>, were unsuccessful, despite trialling three different fluorescence indicators and many assay conditions. The assay proved unreliable and many complex component interactions and factors were encountered that would impede assay optimisation to establish robust activity, making continued efforts unattractive. Initial attempts to establish an alternative EpiQuik™ HAT activity assay were promising and potentially seemed to indicate enzyme activity but there was high associated error.

With hits from the fluorescence based HTS<sup>4,5</sup> likely representing false positives, these were discounted and the available chemical matter for the KAT2A and KAT2B HAT domains was reinvestigated. The only series identified was the isothiazolones, which are well known PAINS and were therefore also disregarded. Unfortunately, this exhausted the inventory of reported small molecule inhibitors of KAT2A and KAT2B and thereby concluded attempts to develop a KAT2A HAT tool inhibitor.

Unfortunately, while HATs represent interesting targets for tool inhibitor development, as understanding of these enzymes improves, numerous challenges are being recognised in how to target them. For example, the HATs have an abundance of cellular substrates and their acetyltransferase activity and specificity for these is often influenced by other proteins. KAT2A is incorporated into multiprotein complexes, which influence the specificity and the catalytic activity towards its histone and non-histone targets.<sup>518</sup> This affect must be addressed in the development of small molecule HAT inhibitors, as the activities of recombinant HAT enzymes may not reflect their *in vivo* activity. This may limit the translation from *in vitro* assays to *in vivo* disease models.

These challenges are not insurmountable. C646 has been reported as a potent and selective active-site directed small molecule inhibitor of p300, Figure 5.21.<sup>519</sup> Structure-based, *in silico* screening was used to identify this commercially available inhibitor. C646 is a competitive p300 inhibitor, with a  $K_i$  of 400 nM and was shown to inhibit histone acetylation and cell growth.<sup>519</sup> Since it was discovered, therapeutic benefits of C646 have been reported in many cancers. For example, it inhibits the growth of human melanoma<sup>520</sup> and induces apoptosis in prostate cancer cells.<sup>521</sup>



C646

**Figure 5.21. Existing p300 HAT Inhibitor C646.** Best validated HAT inhibitor, selective for p300 and CBP.<sup>519</sup>

Despite the associated challenges, novel HAT inhibitors are necessary and highly desirable. The existing KAT2A inhibitors are particularly poorly suited for biological investigation, disease validation and initiating drug discovery efforts. An alternative assay format with robust activity must be identified and a high-throughput screen should be conducted to identify new, reliable chemical matter. Regrettably, this was beyond the scope of this work.



## 6 Design and Characterisation of KAT2A Bromodomain Tool Inhibitors

Despite complications in developing a tool inhibitor for the KAT2A HAT domain, this histone acetyltransferase remained an interesting protein target. KAT2A is synthetic lethal with BAF180<sup>3</sup>, which is frequently mutated in clear cell renal cell carcinoma<sup>376</sup>, and is therefore a compelling target for tool inhibitor development.

In Chapter 4.2, a druggable binding pocket was also identified at the KAT2A bromodomain. Bromodomains are currently a hot topic in research, with many potential applications in diseases, particularly cancers<sup>522</sup>, and therefore represent highly interesting targets for the development of small molecule tool inhibitors. When this work was initiated there were no existing KAT2A or KAT2B inhibitors with sufficient affinity or selectivity to enable biological investigation, disease validation and initiation of drug discovery efforts.

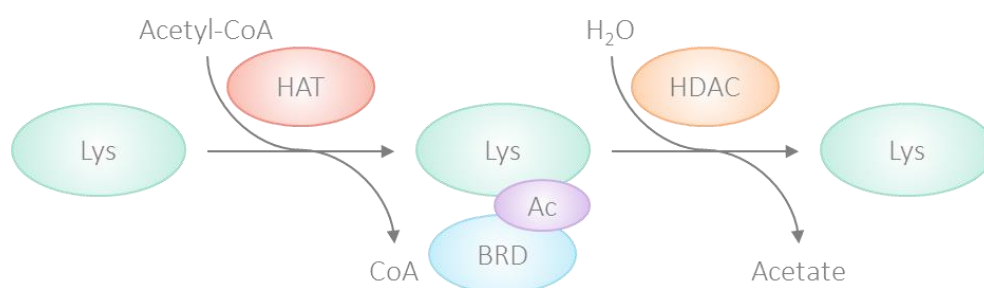
In this chapter, available KAT2A and KAT2B crystal structures were utilised in a computational drug design effort, and T<sub>m</sub> shift, TR-FRET, ITC and X-ray crystallography assay techniques were optimised and established in-house to enable characterisation of prospective tool inhibitors of the KAT2A bromodomain. The assays were employed in screening over 3000 small molecules and fragments to assess their affinity for KAT2A, and an assortment of novel KAT2A bromodomain binders were identified. The optimised assays, novel chemical matter and collection of ligand-bound crystal structures provide an exciting opportunity for further study and should enable development of potent and selective KAT2A bromodomain tool inhibitors.

### 6.1 Introduction

Having encountered a variety of setbacks in developing a tool inhibitor for the KAT2A HAT domain, including exhausting the inventory of small molecule HAT inhibitors of KAT2A and KAT2B, further efforts in developing a HAT tool inhibitor were terminated. Focus was instead redirected to the KAT2A bromodomain. Bromodomains (BRDs) are evolutionarily conserved protein-protein interaction domains, which selectively recognise and bind to acetylated lysine residues, thereby directing protein function.<sup>522</sup> Inhibitors designed to target this domain should disrupt interactions with the acetylated protein targets, which are commonly histones. This will not impede the catalytic activity of KAT2A but should disrupt its function by preventing it from properly localising to target sites. It is hoped that this will afford the desired biological response, to selectively kill BAF180 deficient cells.

### 6.1.1 Function and Phylogeny

Lysine acetylation has emerged as a prevalent regulatory post-translational modification (PTM) in proteins and, as discussed in Chapter 5.1, it is important in transcriptional activation and gene expression. The acetylation of conserved lysine residues located in the amino terminal domains of histones serves to loosen the chromatin structure. Acetylation results in neutralisation of the positive charge of the basic lysines. This weakens the interaction of the histone with the DNA, producing the more open chromatin required for transcription, and establishes binding sites for other proteins and complexes.<sup>433</sup> Deregulation of histone acetylation can induce aberrant expression of critical genes implicated in cell growth, differentiation and apoptosis, and thereby enable proliferation and tumourigenesis.<sup>523</sup> Histone acetyltransferases, such as KAT2A, are responsible for acetylating lysine side chains on histones, by transferring acetyl from acetyl-CoA. These acetyl groups are erased by histone deacetylases. Recognition of acetylated lysine residues is principally mediated by bromodomains, Figure 6.1.

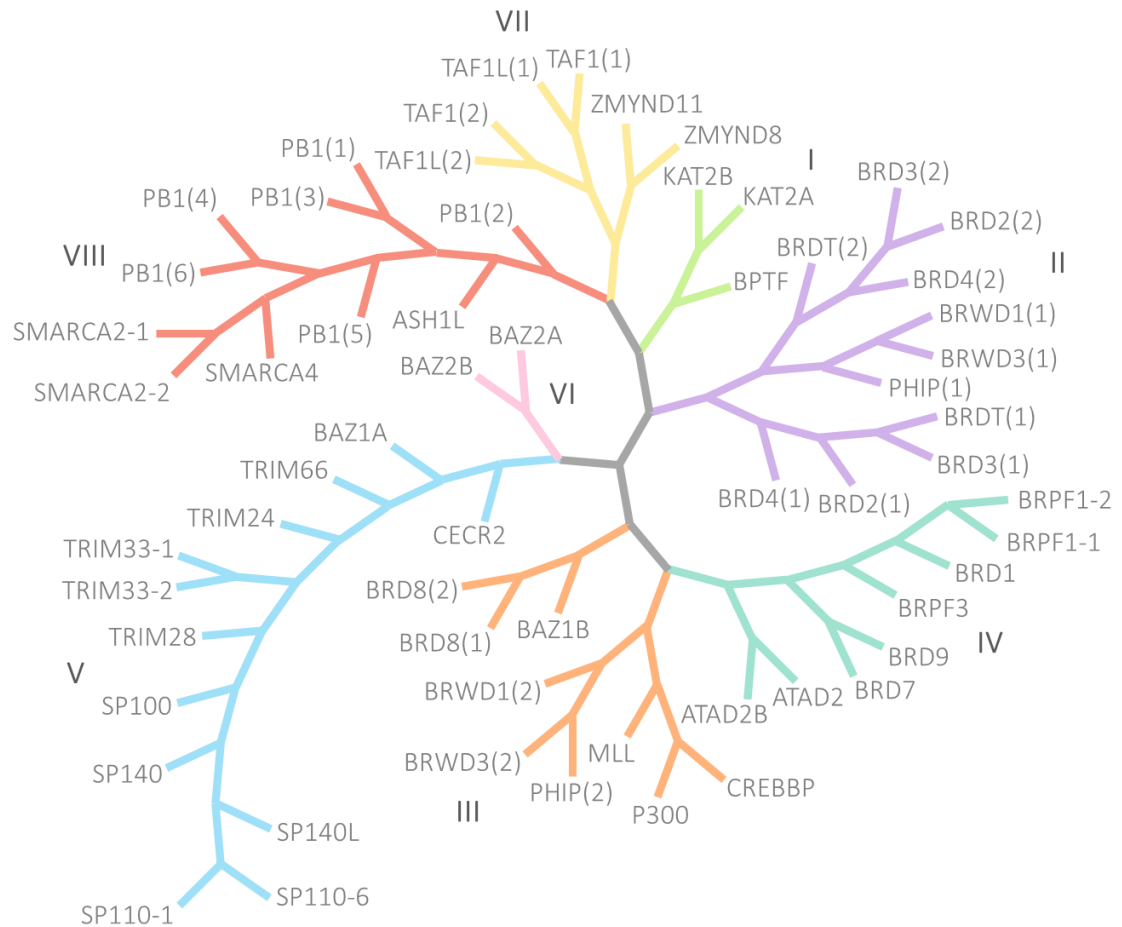


**Figure 6.1. The Mechanism of Histone Acetylation.** Lysine residues are acetylated by HATs, utilising acetyl-CoA, and this is erased by HDACs. BRDs recognise acetylated lysine. Figure adapted from Shortt *et al.*, 2017.<sup>524</sup>

The first BRD was identified in the *Drosophila melanogaster* protein brahma.<sup>525</sup> As this structural motif was conserved in drosophila FSH, yeast SNF2 and SPT7 and human CCG1 and RING3, it was suggested that it may be characteristic of a new family of regulatory proteins.<sup>525,526</sup> Four of these proteins were known to function in activation of other genes, and therefore it was hypothesised that the bromodomain in some way mediates interactions with other proteins that are necessary for transcriptional activation.<sup>525</sup> Secondary structure prediction methods identified amphipathic  $\alpha$ -helices that were nominated as sites of intramolecular or intermolecular protein-protein interaction, which could influence the assembly of multicomponent complexes.<sup>526</sup>

It is now well established that bromodomains selectively recognise and bind to acetylated lysine residues, primarily on histones<sup>522</sup>, thereby acting as epigenetic reader domains.<sup>527</sup> It is therefore perhaps unsurprising that these evolutionarily conserved protein modules are found in many chromatin-associated proteins and the majority of known nuclear HATs.<sup>446</sup> In total, 61 unique BRDs are encoded by the human proteome, in 46 different proteins.<sup>238</sup> As for the HAT domains,

phylogenetic trees have been devised to map the evolutionary and structural relationships of the BRDs, Figure 6.2.<sup>219,380</sup> Common aliases of some of the proteins are listed in Table 6.1.



**Figure 6.2. Phylogenetic Tree of Bromodomains.** Schematic showing bromodomains subdivided into families. Figure adapted from Shortt *et al.*, 2017.<sup>524</sup>

These bromodomains commonly occur in multidomain proteins, in which they are linked via flexible amino acid chains to diverse catalytic domains, such as the HAT domain in KAT2A, or additional epigenetic reader domains, such as methyl-lysine reader PHD (plant homeodomain) fingers, BAH domains or other bromodomains.<sup>238</sup> The linker chain affords conformational flexibility, which enables the protein to interact with diverse motifs. The combining of multiple protein-protein interaction domains is illustrative of a combinatorial recognition process, which further complicates interpretation of the histone code.<sup>522</sup>

Bromodomain-containing proteins have diverse physiological functions and regulate gene expression by modulating transcription via a range of mechanisms. They can function directly as components of chromatin remodelling complexes, perform catalytic functions to modify histones, act as histone-recognising scaffolds to facilitate the assembly of larger protein complexes, or serve as transcriptional co-regulators, Figure 6.3.<sup>522</sup>

**Table 6.1. Bromodomains.** List of common aliases of subset of the bromodomains.

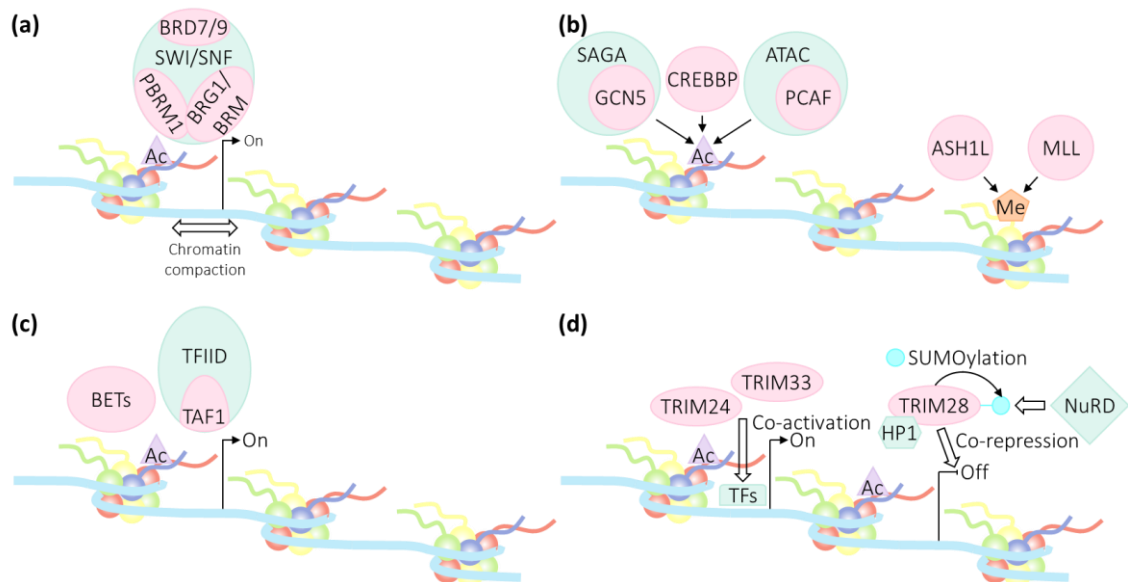
Given Name	Alias	Given Name	Alias
KAT2A	GCN5	BAZ2A	TIP5
KAT2B	PCAF	TRIM24	TIF1 $\alpha$
BPTF	FALZ, FAC1	TRIM28	TIF1 $\beta$ , KAP1
BRD2	RING3	TRIM33	TIF1 $\gamma$
BRWD1	WDR9	SP140L	LOC93349
PHIP	WDR11	SMARCA2	BRM, BAF190B, SNF2A
BRD1	BRPF2	SMARCA4	BRG1, BAF190A, SNF2B
ATAD2B	KIAA1240	PB1	PBRM1, BAF180
CREBBP	CBP, KAT3A	ASH1L	KMT2H
P300	KAT3B, EP300	TAF1	KAT4, TAFII250, TFIID1, CCG1
MLL	KMT2A	ZMYND8	PRKCBP1, PKCB1
BAZ1B	WSTF		

The multisubunit BAF and PBAF complexes, which have key roles in cell differentiation and proliferation<sup>393,397</sup> and DNA repair<sup>399</sup>, contain multiple BRD-containing proteins. The central ATPases, BRG1 and BRM, are responsible for ATP-dependent alteration of chromatin structure, and contain carboxy-terminal BRD modules.<sup>238</sup> Closely related proteins BRD9 and BRD7, specific to BAF and PBAF respectively, contain individual bromodomains. Additionally, PBAF contains a third bromodomain-containing protein, PBRM1, which contains six tandem BRDs, two bromo-adjacent homology (BAH) domains, and a high-mobility group (HMG).<sup>376</sup>

Catalytic activity is observed in several bromodomain-containing proteins, including PCAF, GCN5, CREBBP, ASH1L (absent, small and homeotic disks protein 1 homologue) and MLL (mixed-lineage leukaemia). These proteins modify chromatin structure by introducing PTMs on histones. The bromodomains bestow specificity by tethering the proteins to particular chromosomal sites, and thereby modulate transcriptional programs and drive phenotypic changes.<sup>446</sup> PCAF, GCN5 and CREBBP have catalytic acetyltransferase activity and thereby regulate transcriptional activation.<sup>437,528</sup> ASH1L and MLL have catalytic methyltransferase activity, which can cause relaxation or compaction of chromatin. ASH1L contains BRD, PHD finger and BAH domains as well as the SET (Su(var)3-9, Enhancer-of-zeste and Trithorax) domain, which is responsible for its methyltransferase activity.<sup>529</sup> MLL contains BRD and SET domains and three PHD fingers.<sup>530</sup>

All bromodomain-containing proteins are able to recognise acetylated lysine residues, but for some, which serve to recruit transcriptional regulators or machinery to specific loci, this is their primary function. The most prominent example is the BET proteins, BRD2, BRD3, BRD4 and testis-specific BRDT. These proteins harbour tandem, amino-terminal bromodomains, which bind to acetylated lysines, and an extra-terminal (ET) protein-protein interaction domain, which

interacts with specific transcriptional effectors.<sup>531</sup> Utilising these interaction domains, the BET proteins act as scaffolds to recruit transcriptional machinery components to conserved loci and thereby regulate transcriptional activity. A second example is TAF1, which is the largest component of the TFIID transcription factor complex. TAF1 also contains tandem bromodomains that bind selectively to multiply acetylated histone H4 peptides.<sup>532</sup> It modulates transcriptional initiation by binding at the promoter region and recruiting other transcriptional regulators.<sup>533</sup>



**Figure 6.3. Roles of BRD-Containing Proteins in Regulating Gene Expression.** Schematic detailing range of mechanisms by which bromodomain-containing proteins can modulate transcription. (a) BRG1, BRM, BRD7, BRD9 and PBRM1 are components of SWI/SNF chromatin remodelling complexes. (b) GCN5, PCAF and CREBBP exhibit catalytic histone acetyltransferase activity and thus regulate transcriptional activation. ASH1L and MLL exhibit catalytic histone methyltransferase activity, which can relax or compact chromatin. (c) The BET proteins and TAF1 act as histone-recognising scaffolds, which recruit transcriptional effectors. (d) The TRIM family proteins function as co-regulators to modulate transcriptional activity. Figure adapted from Fujisawa & Filippakopoulos, 2017.<sup>522</sup>

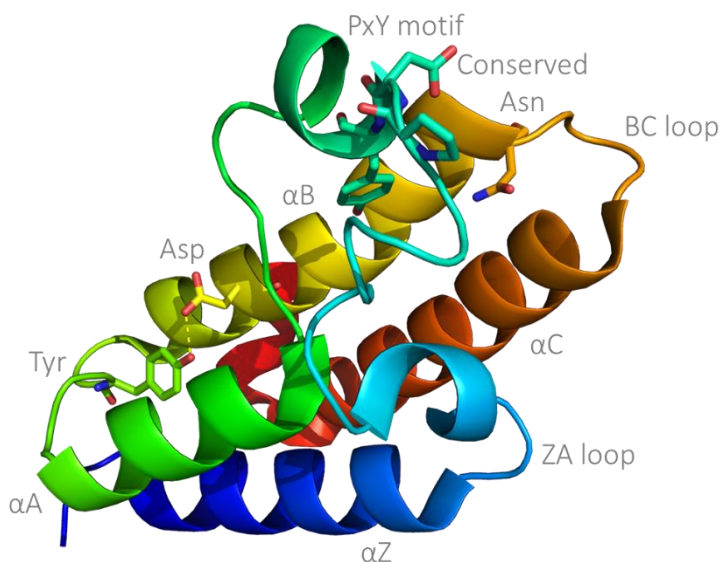
Finally, some BRD-containing proteins, such as the TRIM family members, act as transcriptional co-regulators. TRIM24 is a co-activator of oestrogen receptor  $\alpha$ , a ligand-activated transcription factor.<sup>534</sup> TRIM28 has roles in transcriptional repression by modulating chromatin structure via interaction with heterochromatin protein 1 (HP1)<sup>535</sup>, and mediating intramolecular SUMOylation, which provokes gene silencing via recruitment of the NuRD (nucleosome remodelling and deacetylase) complex.<sup>536</sup> TRIM33 exhibits E3 ubiquitin ligase activity and regulates the TGF- $\beta$  signalling pathway, which induces transcription of a variety of integral genes.<sup>537</sup>

Large-scale studies have identified mutations in various bromodomain-containing proteins that lead to cancer. For example, an analysis of 21 paediatric cancer subtypes identified recurrent mutations in 21 BRD-containing proteins.<sup>538</sup> No systematic studies have yet determined the mutation load within the BRD modules however, and therefore it is not clear what contribution acetyl-lysine readout makes towards these phenotypes.<sup>522</sup> This is a common obstacle in the

study of multidomain proteins. Specific bromodomain tool inhibitors will aid in elucidating the functions of these modules in physiology and disease.

### 6.1.2 Structure and Interactions

High-resolution X-ray or NMR structures are now available for 53 of the 61 BRDs<sup>219</sup>, thanks in large part to the Structural Genomics Consortium. This abundance of structural information has enabled the comprehensive structural analysis of bromodomains, to establish the sequence and structural conservation of the modules.<sup>539</sup> Bromodomains are small modules of approximately 110 amino acids<sup>540</sup>, which despite a low degree of overall sequence homology, share a conserved architecture. The globular fold, which was first characterised almost twenty years ago<sup>541</sup>, comprises of a left-handed bundle of four  $\alpha$ -helices ( $\alpha$ Z,  $\alpha$ A,  $\alpha$ B and  $\alpha$ C), linked by two variable loop regions (ZA and BC loops), Figure 6.4.<sup>238</sup> The loop regions form the binding site for acetylated lysine residues.<sup>446</sup>



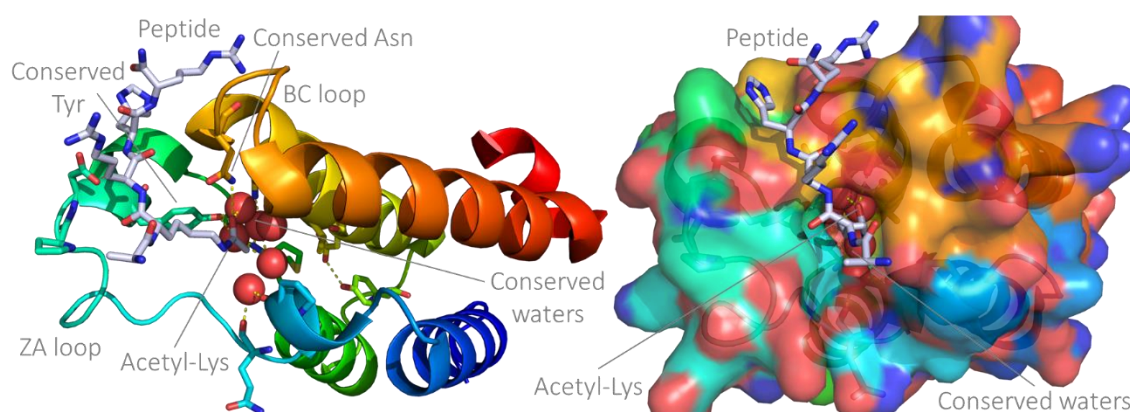
**Figure 6.4. Conserved Structure of Bromodomain Modules.** Crystal structure of *Saccharomyces cerevisiae* GCN5 bromodomain (PDB ID: 1E6I)<sup>239</sup>, with protein chain coloured from blue at N-terminus to red at C-terminus.

The four helices form a deep cavity, flanked by the ZA and BC loops, to create a surface-accessible hydrophobic binding pocket, which accommodates the neutralised acetyl-lysine.<sup>238</sup> This pocket is surrounded by a charged surface that interacts with the backbone of the bound peptide.<sup>522</sup> In contrast to the conserved fold, these protein-interaction surfaces are highly diverse, exhibiting electrostatic potentials ranging from strongly positively to strongly negatively charged.<sup>238</sup> This indicates that different bromodomains recognise distinct peptide sequences. The BRDs with highly positively charged interaction surfaces are intriguing as histone peptides are typically rich in positively charged lysine and arginine residues and will therefore interact poorly with these. Given the numerous bromodomain-containing proteins, it is plausible that an array of other acetylated protein targets might exist besides the histones.<sup>522</sup> So far however,

only a few examples are known, such as the interaction of PCAF with HIV-1 protein Tat acetylated at lysine 50<sup>542</sup>, or CREBBP with p53 acetylated at C-terminal lysine 382.<sup>543</sup>

In 2012, Filippakopoulos *et al.* conducted a comprehensive study of BRD structure and peptide binding properties, in attempt to better understand acetylation-mediated signalling, as interpreted by bromodomains.<sup>238</sup> They crystallised and determined the structures of 29 BRDs, including 25 novel structures, and performed a SPOT blot analysis to assess the ability of 33 representative human bromodomains to recognise acetylated histone peptides spanning known acetyl-lysine sites.<sup>544</sup> In general, they found that the binding of bromodomains to acetylated histone peptides is relatively weak, indicative of additional interaction domains *in vivo*, and that peptide recognition is typically dependent on patterns of multiple PTMs rather than single acetylation marks. They also uncovered that of the acetyl-lysine sites within the histone tails, H2AK15ac, H2AK36ac, H3K14ac and H4K5ac were the most commonly recognised by BRDs.<sup>238</sup> The study provides a powerful resource to enable future functional studies of these epigenetic reader domains.

Despite their divergence, the bromodomains contain several highly conserved residues and motifs, Figure 6.4. Two residues understood to facilitate the interaction with acetyl-lysine are observed in the majority of BRDs. An asparagine within the BC loop forms a direct hydrogen bond with the acetyl group, while a tyrosine in a conserved PxY motif in the ZA loop interacts with the lysine residue via a water-mediated hydrogen bond, Figure 6.5. Other conserved motifs are believed to stabilise the structural fold. For example, a tyrosine residue in the AB loop forms a charge-assisted hydrogen bond with an aspartic acid residue in  $\alpha$ B, Figure 6.4.



**Figure 6.5. Interaction of Bromodomains with Acetylated Peptides.** Crystal structure of *Saccharomyces cerevisiae* GCN5 bromodomain complexed with acetylated H4 peptide (PDB ID: 1E6I).<sup>239</sup> Protein chain coloured from blue at N-terminus to red at C-terminus and conserved structural elements annotated.

As discussed, there are now many bromodomain crystal structures available, many of which depict them in complex with acetylated histone peptides. The structures confirm the mode of binding of acetyl-lysine. As well as the conserved asparagine and tyrosine, an array of water

molecules is invariably observed at the bottom of the binding cavity, which interact in a consistent manner with histone peptides and other ligands, Figure 6.5.<sup>545</sup> It is important that these conserved waters are considered in BRD drug design and virtual screening efforts.<sup>539</sup>

Some BRD-containing proteins are able to interact with multiple acetyl-lysines. Predictably, proteins with tandem bromodomains joined by a flexible linker, such as the BET proteins and TAF1, are able to simultaneously recognise two distinct acetylated lysine residues that exist on the same or on different proteins.<sup>546</sup> This increases their specificity for multiply acetylated histone tails.<sup>532</sup> Furthermore, some single bromodomains are able to recognise diacetylated peptides. For example, murine BRDT interacts with histone H4 acetylated at K5 and K8. Both acetyl-lysine residues are accommodated in the same hydrophobic binding cavity.<sup>547</sup> It seems that this ability to recognise diacetylated histone peptides is a common feature of the BET bromodomains, which exhibit enlarged binding pockets.<sup>238</sup>

In Chapter 4.2, a druggable binding pocket was identified in the KAT2A bromodomain, which coincides with the binding site of acetylated histone peptides, composed of the hydrophobic acetyl-lysine binding pocket and the surrounding charged surface. This indicates that this binding site could be susceptible to targeting with a small molecule inhibitor. The high sequential and structural variability at the comprising ZA and BC loop regions should enable discrimination of different bromodomain targets<sup>548</sup> and thereby afford selectivity potential, making this an attractive target site for inhibitor development.

### 6.1.3 Existing Inhibitors

Bromodomains have emerged as interesting targets for the development of protein interaction inhibitors and currently represent a particularly lively area of research. This is in part enabled by scientists at the SGC, who aim to develop chemical probes and optimise therapeutic leads to encourage translation of small-molecule modulators of epigenetic targets to therapeutics. An illustrative subset of the BRD inhibitors reported to date is reviewed below.

The first small molecule BRD inhibitors were reported by the Zhou laboratory. Firstly, in 2005, using a combination of NMR-based chemical screening and an ELISA assay, N1-aryl-propane-1,3-diamine compounds were identified as a novel class of PCAF BRD inhibitors, capable of displacing acetylated HIV-1 transactivator Tat peptide (Tat-Ack50) from KAT2B.<sup>548</sup> In 2007, structure-guided optimisation was used to determine a related small molecule chemical inhibitor, Figure 6.6, which exhibited an  $EC_{50}$  of  $\sim 1 \mu M$  on inhibiting Tat-mediated transcription of the viral promoter in a HIV-1 long terminal repeat-luciferase reporter gene assay.<sup>549</sup> These ligands purportedly bind the KAT2B BRD selectively over structurally similar CREBBP and TRIM28 BRDs,



capitalising on the aforementioned sequence variations in the ZA and BC loops to discriminate between binding targets.<sup>548</sup> Secondly, in 2006, a similar NMR-based approach was used to identify small molecules that block lysine 382-acetylated p53 from interacting with CREBBP, from a focused chemical library generated based on structural knowledge of the CREBBP BRD-p53-Ack382 interaction.<sup>550</sup> Hits were confirmed using structure-based characterisation, an *in vitro* protein-peptide binding assay, cell-based functional assays and fluorescence spectroscopy, in which the most potent inhibitor, MS7972, Figure 6.6, exhibited a  $K_d$  of  $\sim 20 \mu\text{M}$ .<sup>550</sup>

2010 saw the landmark discoveries of potent BET BRD inhibitors, JQ1 and I-BET762, Figure 6.6, which bind competitively with histones at the acetyl-lysine recognition motif, Figure 6.7.<sup>217,218</sup> These triazolodiazepines are selective for the BET family but not individual members. Developing inhibitors specific for a single BET bromodomain will aid in elucidating the functional roles of the proteins and remains an important challenge for the field.<sup>545</sup> JQ1 exhibited  $K_d$ s of  $\sim 50 \text{ nM}$  for BRD4(1),  $\sim 90 \text{ nM}$  for BRD4(2) and  $60\text{--}190 \text{ nM}$  for the other family members, and showed preclinical efficacy as a NUT (nuclear protein in testis) midline carcinoma therapeutic, targeting the BRD4-NUT fusion oncoprotein.<sup>217</sup> The triazole moiety mimics the acetyl-lysine, forming hydrogen bonds with the signature asparagine and conserved waters. I-BET762 was developed and investigated independently but simultaneously. It presents interactions, binding affinities and selectivity profile with BET bromodomains similar to JQ1.<sup>218,551</sup> The publication of these potent BET inhibitors with translational potential markedly elevated interest in BRD inhibitors.

As discussed, although all bromodomains contain the hydrophobic acetyl-lysine binding pocket, the variable residues at the ZA and BC loop region result in diversity in the environments around that small cavity. As such, druggability varies considerably for different bromodomains, Figure 6.8.<sup>552</sup> Currently, the majority of existing BRD inhibitors target the highly druggable BET proteins. Like JQ1 and I-BET762, many of these BRD inhibitors have rigid aromatic scaffolds and present hydrogen bond acceptors to mimic the acetyl-lysine, Figure 6.7.<sup>553</sup> Triazolodiazepine, OTX015<sup>554</sup>, Figure 6.6, derived from JQ1, is currently in clinical trials for various malignancies including lymphoma and leukaemia.<sup>555,556</sup> The structural rigidity and hydrophobicity of these diazepines confers good shape complementarity and specificity for the BET bromodomains but limits bioavailability and chemical diversity.<sup>545</sup> The more adaptable dimethylisoxazole scaffold was discovered fortuitously by Hewings *et al.*<sup>557</sup> following the identification of DMSO (dimethyl sulfoxide) and NMP (N-Methyl-2-pyrrolidone) as BRD ligands.<sup>558</sup> A number of dimethylisoxazole BET inhibitors have now been developed, including I-BET151, Figure 6.6, which exhibits efficacy against human and murine MLL-fusion leukaemic cell lines.<sup>559</sup> I-BET151 retains the potency and selectivity of I-BET762 while affording improved pharmacokinetics and half-life.

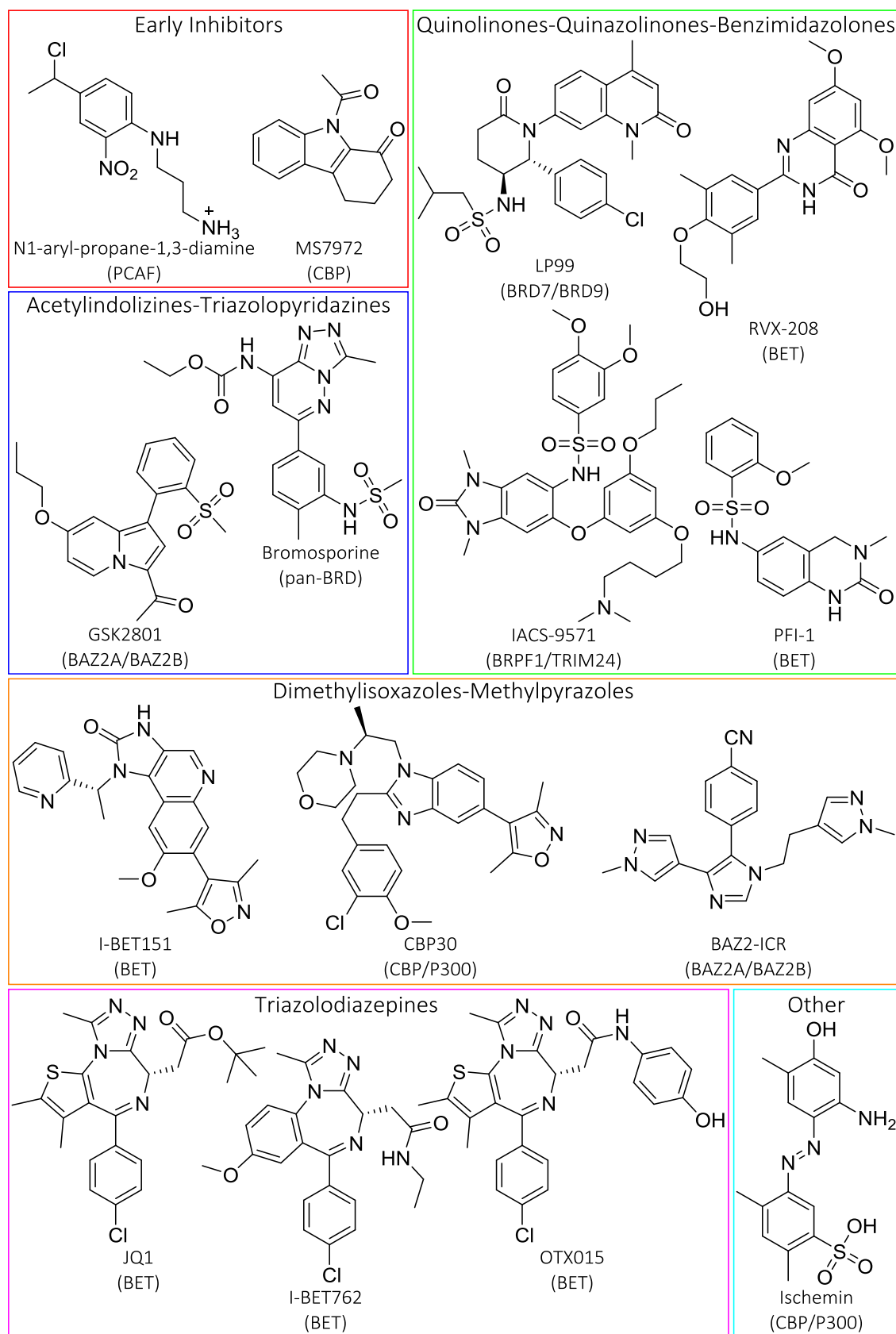
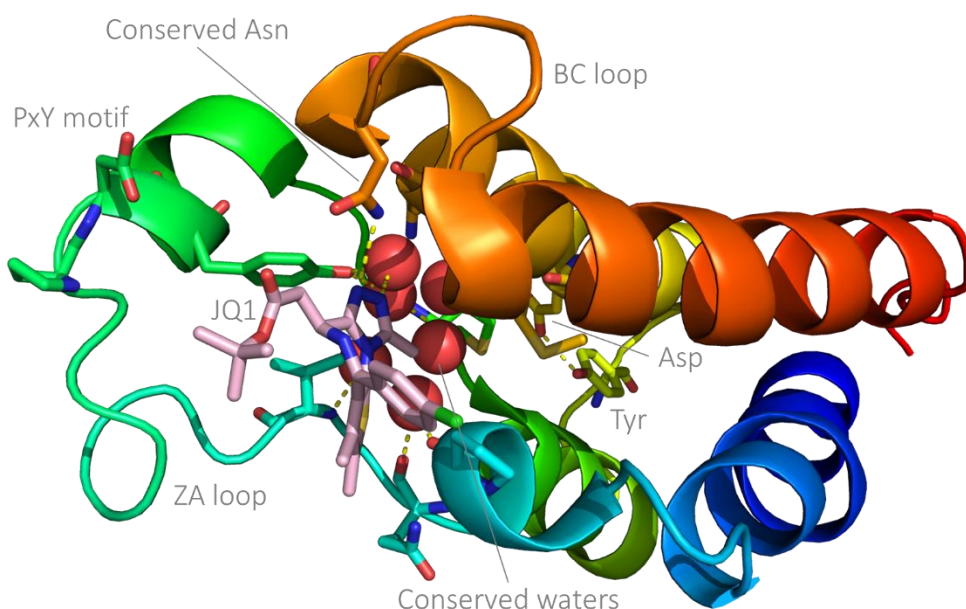


Figure 6.6. Existing Bromodomain Inhibitors.

As well as the diazepines and dimethylisoxazoles, a number of other scaffolds have been successfully employed in BET inhibitors. Quinazolinones were identified as effective inhibitor

scaffolds via fragment-based screening and optimised to afford the potent, BET-selective inhibitor, PFI-1, Figure 6.6.<sup>560</sup> PFI-1 has antiproliferative effects on leukaemic cell lines exhibiting MLL rearrangements.<sup>561</sup> RVX-208, Figure 6.6, a quinazolinone derivative of resveratrol, was discovered as a BET BRD inhibitor in 2013, and is currently the sole BET inhibitor that offers any selectivity for one of the tandem BRDs (BRD3(2)  $K_d$  = 195 nM, BRD3(1)  $K_d$  = 4.1  $\mu$ M).<sup>562</sup>

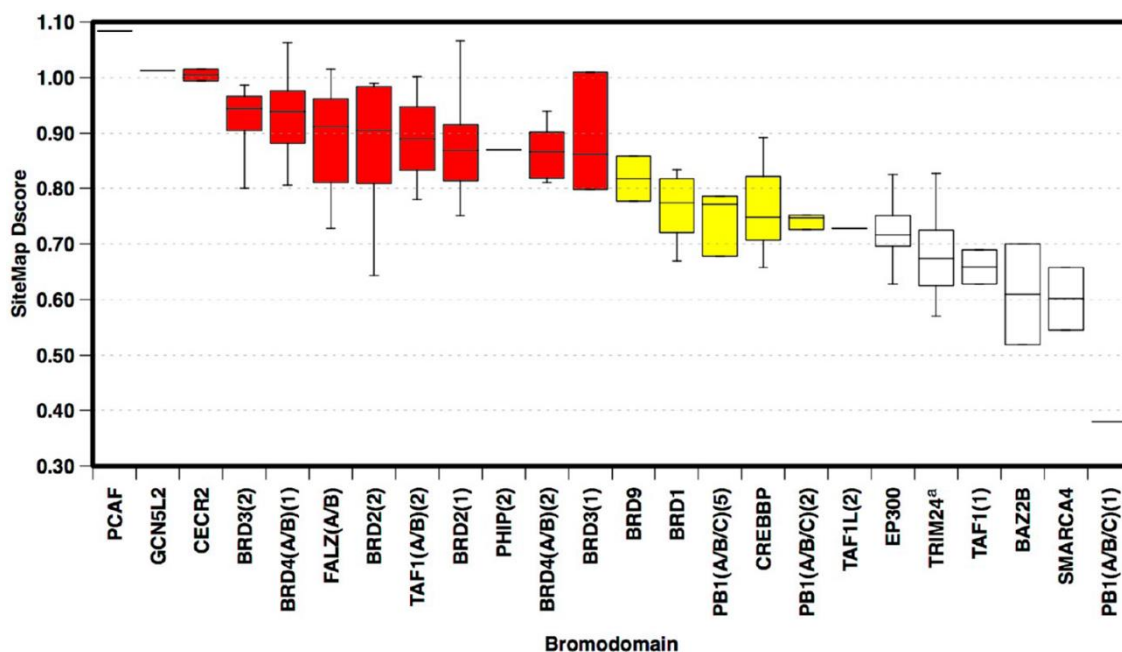


**Figure 6.7. Interaction of BRD4 with JQ1.** Crystal structure of *Homo sapiens* BRD4 bromodomain in complex with inhibitor JQ1 in pink (PDB ID: 3MXF).<sup>217</sup> Protein chain coloured from blue at N-terminus to red at C-terminus.

Closely related CREBBP and p300 are theoretically less druggable than the BET proteins, Figure 6.8.<sup>552</sup> Nevertheless, likely inspired by the success of MS7972<sup>550</sup>, a number of research groups have ventured to develop potent and selective CREBBP/p300 inhibitors. Ischemin, Figure 6.6, was developed using a structure-guided approach from hits from an NMR-based screen of 3000 compounds.<sup>563</sup> Ischemin inhibited CREBBP, exhibiting a  $K_d$  of 19  $\mu$ M, and was observed to down-regulate p53-induced apoptosis. The dimethylisoxazole scaffold was also discovered to be effective for CREBBP inhibition. A structure-guided approach was utilised to develop a potent and selective inhibitor of CREBBP and p300. CBP30, Figure 6.6, exhibits  $K_d$ s of 21 nM and 32 nM for CREBBP and p300 respectively, and 40-fold selectivity over the BET bromodomains.<sup>564</sup>

Despite them being classified among the least druggable bromodomains due to their shallow binding pockets<sup>552</sup>, inhibitors have also been reported for the bromodomain adjacent to zinc finger domain (BAZ) proteins, BAZ2A and BAZ2B. In 2015, the SGC utilised virtual screening and structure-guided design to develop methylpyrazole BAZ2-ICR, Figure 6.6.<sup>565</sup> BAZ2-ICR binds to BAZ2A and BAZ2B with  $K_d$ s of 109 nM and 170 nM respectively, and is selective against all other BRDs with the exception of CECR2 (Cat Eye syndrome chromosome region, candidate 2), which it binds with a  $K_d$  of 1.55  $\mu$ M. In collaboration with GlaxoSmithKline, the SGC also developed

acetylindolizine GSK2801, Figure 6.6, as a potent and reasonably selective inhibitor of BAZ2A and BAZ2B.<sup>566</sup> GSK2801 offers  $K_d$ s of 257 nM and 136 nM for BAZ2A and BAZ2B respectively, but also binds to TAF1L ( $K_d$  = 3.2  $\mu$ M) and BRD9 ( $K_d$  = 1.1  $\mu$ M), albeit more weakly.



**Figure 6.8. Box Plots Showing Range and Distribution of BRD Druggability Assessed by SiteMap.** Domains ranked by median Dscore and coloured to indicate druggability classification: red, druggable; yellow, intermediate; white, difficult. Figure from Vidler *et al.* (2012).<sup>552</sup>

Dual inhibitors of BRPF1 (bromodomain and PHD finger containing 1) and TRIM24 have also been developed. For example, in 2015, benzimidazolone IACS-9571, Figure 6.6, was reported as a potent inhibitor with  $K_d$ s of 31 nM and 14 nM for TRIM24 and BRPF1 respectively.<sup>567</sup> IACS-9571 has yet to be validated in cell or animal disease models.<sup>545</sup>

Finally, in 2015, the quinolinone-fused lactam LP99, Figure 6.6, was reported as the first potent, selective and cell-active inhibitor of the BRD7 and BRD9 bromodomains.<sup>568</sup> LP99 was developed from a fragment hit via an approach that combined structure-based inhibitor design and biophysical characterisation with tractable chemical synthesis.

Potent and selective inhibitors afford new therapeutic strategies and are crucial for biological exploration of the target protein.<sup>527</sup> As a result, publication of a new chemical probe often provokes an increased level of research in the area. The past five years have therefore seen an explosion in research into the therapeutic application of BRD inhibitors, which has facilitated a recent deluge of BET inhibitors entering clinical trials as anticancer agents.<sup>569</sup>

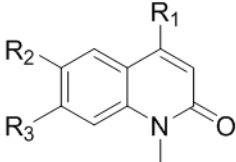
Despite the considerable efforts of the SGC and other academic groups, selective inhibitors are not yet described for every bromodomain target. As such, tool compounds with a wider specificity profile are also useful in research, for example in elucidating the universal roles of

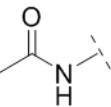
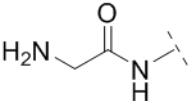
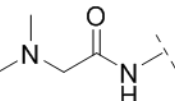
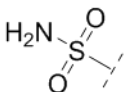
bromodomains and in establishing functional *in vitro* and cellular assays. Bromosporine, Figure 6.6, is a pan-BRD inhibitor described by the SGC with activity against the BET, CECR2, TAF1, BRD9 and CREBBP bromodomains (<http://www.thesgc.org/chemical-probes/bromosporine>).

#### 6.1.4 Recent KAT2A/B Chemical Matter

Despite the early report of N1-aryl-propane-1,3-diamine compounds as potential PCAF BRD inhibitors<sup>548</sup>, and GCN5 and PCAF representing two of the most highly druggable bromodomains, Figure 6.8<sup>552</sup>, when this work was initiated there were no potent and selective inhibitors reported for either of these closely related homologues. However, the SGC had utilised thermal stability shift assays, ITC and high-throughput X-ray crystallography to identify fragments with inhibitory activity against the PCAF bromodomain, Table 6.2, which they shared with us pre-publication.<sup>240</sup>

**Table 6.2. Available Biophysical Data for Novel Fragment Inhibitors of KAT2B/PCAF Bromodomain.**<sup>240</sup>



Compound	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	Crystal Structure	T <sub>m</sub> Shift at 1 mM (°C)	ITC K <sub>d</sub> (μM)
BR004	H	H	H	Y	4.5	320
SL1121	CH <sub>3</sub>	H	NH <sub>2</sub>	N	5.1	160
SL1126	CH <sub>3</sub>	H		Y	5.7	30
SL1134	CH <sub>3</sub>	H		N	3.0	30
SL1132	CH <sub>3</sub>	H		N	4.2	35
SL1136	CH <sub>3</sub>	NH <sub>2</sub>	H	N	3.8	356
SL1131	CH <sub>3</sub>		H	N	2.0	5000

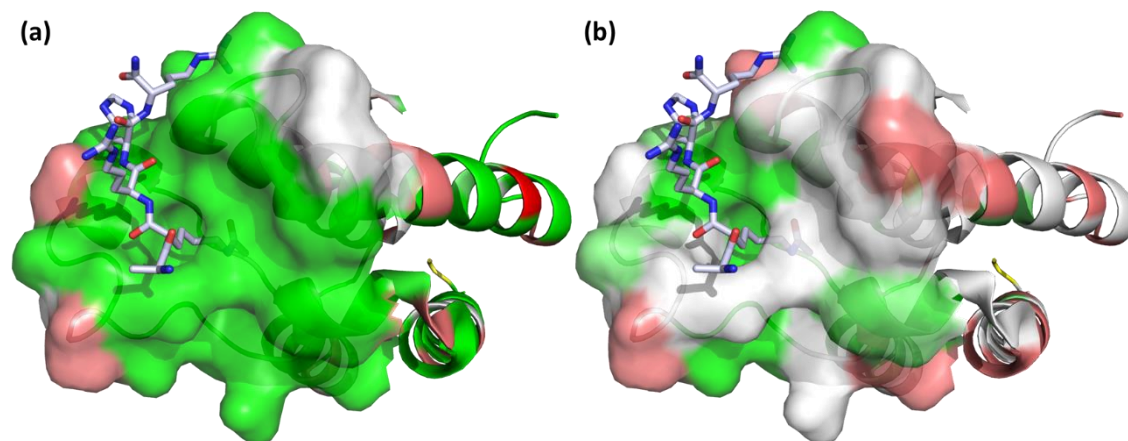
Of these fragments, SL1126 represented the most promising start point for structure-based design, with reasonable affinity (K<sub>d</sub> = 30 μM) and an available X-ray crystal structure (PDB ID: 5FE8).<sup>240</sup> Notably, the fragments showed no selectivity for the KAT2B BRD over the other bromodomains. In fact, many demonstrated higher affinity for other bromodomains, for example BR004 has a K<sub>d</sub> of 5.0 μM for BRD9.<sup>568</sup> With this in mind, strategies to develop a potent

and selective small molecule inhibitor of the KAT2A bromodomain, with low micromolar or nanomolar affinity, should incorporate both growth and diversification of the existing quinolinone scaffold and screening to identify other potential cores.

## 6.2 Results

### 6.2.1 Selectivity Potential

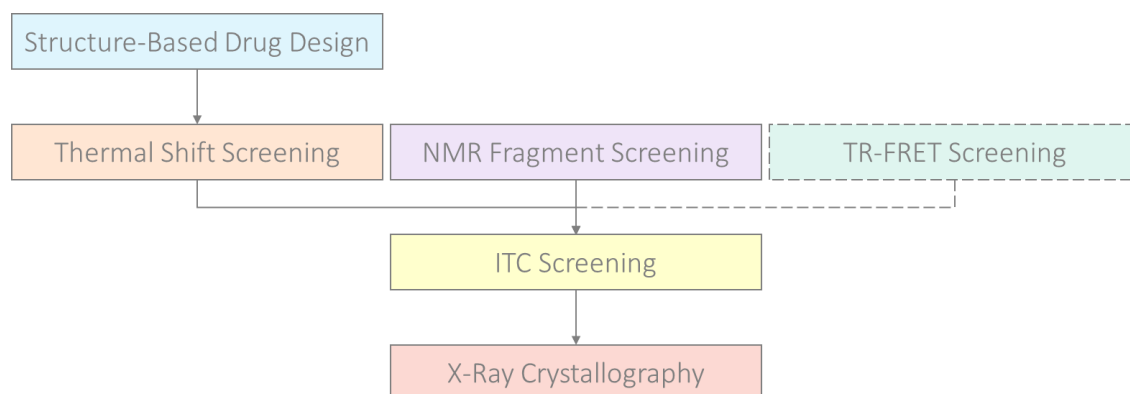
In Chapter 4.2, a druggable structural pocket was identified at the KAT2A BRD using Fpocket 2.0<sup>165,170</sup>, which corresponds to the binding position of acetyl-lysine. As at the HAT domain, EMBL-EBI Clustal Omega<sup>181,182</sup> was used to compile pairwise and multiple sequence alignments to assess the potential for selectivity at this site, Figure Apx.6. Those amino acids which comprise the KAT2A BRD were identified and Scorecons<sup>220</sup> was used to rate sequence conservation. As seen at the HAT domain, the KAT2A and KAT2B BRDs exhibit high homology, Figure 6.9 (a), but the residue conservation between bromodomain families is poorer, Figure 6.9 (b). It is therefore likely that dual inhibitors, which are selective for the two GCN5 homologues but not able to distinguish between them, can be achieved. This was deemed acceptable, especially considering the huge impact of JQ1, which targets all members of the BET BRD family.<sup>217</sup> Small molecule dual inhibitors will serve as useful tools to enable biological exploration of KAT2A and KAT2B to further elucidate their biological roles, as well as aid in establishing assays.



**Figure 6.9. Assessing Residue Conservation at KAT2A Bromodomain.** *Homo sapiens* GCN5 bromodomain (PDB ID: 3D7C)<sup>238</sup> aligned with that of *S. cerevisiae* (PDB ID: 1E6I)<sup>239</sup>, to show acetyl-lysine binding site. Histone H4 peptide shown in grey. KAT2A coloured to indicate amino acid conservation determined by Scorecons<sup>220</sup>; red indicates the least conserved regions and bright green the most. (a) Results from pairwise alignment of KAT2A and KAT2B. (b) Results from multiple sequence alignment of the type I and II bromodomains.

### 6.2.2 Screening Cascade

The KAT2A bromodomain therefore represents a desirable and promising target for development of a potent and specific tool inhibitor. A screening cascade was designed, incorporating the screening techniques routinely adopted by the SGC and other research groups in identifying bromodomain tool inhibitors, Figure 6.10.<sup>217,240,568</sup>



**Figure 6.10. Schematic of Proposed KAT2A BRD Screening Cascade.**

It was proposed that a structure-based approach would be utilised to design a library of potential KAT2A BRD inhibitors based on the quinolinone scaffold of SL1126. This library of inhibitors would be screened using a thermal stability shift assay to identify those with affinity for KAT2A. In tandem, an NMR fragment screen would be conducted to identify novel scaffolds for inhibitor development. Hits from the thermal shift and NMR screens would be confirmed using ITC and crystallographic techniques.

In addition, it was decided that once screening with thermal shift, NMR and ITC was underway, a time-resolved fluorescence resonance energy transfer (TR-FRET) assay would be developed in-house for use in high-throughput screening. While thermal shift assays are straightforward to establish and offer the capacity for good throughput, they can be limited by poor sensitivity and high frequency of false positives. TR-FRET technology has not yet been used to screen for KAT2A BRD inhibitors, but has been utilised with other bromodomains, for example BRD4(1) and CREBBP and p300.<sup>570,571</sup>

### 6.2.3 Structure-Based Design From SL1126

As discussed, fragments had been identified by the SGC that exhibited micromolar affinity for the KAT2B bromodomain, and of these, SL1126 represented the most promising start point for a structure-based design approach.<sup>240</sup> To commence this structure-based drug design, EMBL-EBI Clustal Omega<sup>181,182</sup> was employed to compile a multiple amino acid sequence alignment of the 61 unique human bromodomains, Figure Apx.7, with an aim to identify residues which differ throughout the bromodomains and therefore should enable selectivity towards KAT2A and KAT2B. Ten of the type V BRDs as well as MLL and ZMYND11 (Zinc finger Myeloid, Nervy, and DEAF-1 domain-containing protein 11), exhibited significant divergence from the other bromodomains and consequently aligned poorly. This substantial diversity should make it easy to select against these proteins and therefore they were excluded from subsequent analysis.



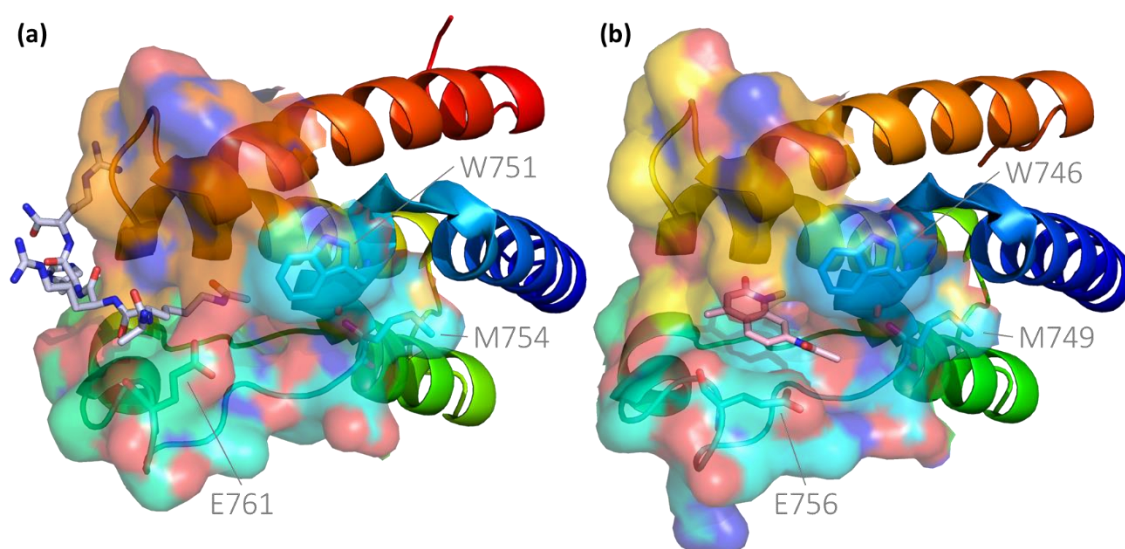
**Table 6.3. Comparison of Bromodomain Binding Site Residues.** Multiple sequence alignment generated by EMBL-EBI Clustal Omega.<sup>181,182</sup> Residues comprising druggable pocket at the acetyl-lysine binding site selected by consulting PDB structures of KAT2A (3D7C)<sup>238</sup> and KAT2B (5FE8)<sup>240</sup>. Conservation scores determined using Scorecons<sup>220</sup> and amino acid residues coloured accordingly; pink indicates the least conserved regions and dark green the most.

BRD	Start Position	1	2	3	4	5	6	7	8	11	12	13	14	15	54	57	58	64
KAT2A	751	W	P	F	M	E	P	V	K	E	A	P	D	Y	C	Y	N	Y
KAT2B	746	W	P	F	M	E	P	V	K	E	A	P	G	Y	C	Y	N	Y
BPTF	2950	W	P	F	L	E	P	V	D	D	A	P	D	Y	C	Y	N	F
BRDT(1)	50	W	P	F	Q	R	P	V	D	Q*	L*	P*	D*	Y*	C*	Y*	N*	I*
BRD4(1)	81	W	P	F	Q	Q	P	V	D	N*	L*	P*	D*	Y*	C*	Y*	N*	I*
BRD3(1)	57	W	P	F	Y	Q	P	V	D	N*	L*	P*	D*	Y*	C*	Y*	N*	I*
BRD2(1)	97	W	P	F	R	Q	P	V	D	G*	L*	P*	D*	Y*	C*	Y*	N*	I*
BRDT(2)	293	W	P	F	Y	N	P	V	D	G*	L*	H*	N*	Y*	C*	Y*	N*	V*
BRD4(2)	374	W	P	F	Y	K	P	V	D	G*	L*	H*	D*	Y*	C*	Y*	N*	V*
BRD3(2)	332	W	P	F	Y	K	P	V	D	E*	L*	H*	D*	Y*	C*	Y*	N*	V*
BRD2(2)	370	W	P	F	Y	K	P	V	D	G*	L*	H*	D*	Y*	C*	Y*	N*	V*
PHIP(1)	1182	S	A	F	V	A	P	V	D	A	Y	P	M	Y	T	F	N	I
BRWD1(1)	1183	A	A	F	A	G	P	V	D	T	Y	P	K	Y	A	F	N	I
BRWD3(1)	1164	S	P	F	A	V	P	V	D	A	Y	P	L	Y	A	F	N	I
BRD8(1)	730	N	V	F	L	Q	P	V	T	I	A	P	G	Y	A	Y	N	V
BRD8(2)	1126	S	P	F	L	K	P	V	S	Q	A	P	G	Y	A	Y	N	V
BAZ1B	1362	W	P	F	R	E	P	V	T	E	A	E	D	Y	A	Y	N	V
BRWD3(2)	1323	E	P	F	R	Q	P	A	D	S	Y	P	D*	Y*	S*	Y*	T*	I**
BRWD1(2)	1336	E	P	F	R	Q	P	V	D	E	Y	P	D	Y	A	Y	T	I
PHIP(2)	1339	E	P	F	R	Q	P	V	D	E	Y	P	D	Y	S	Y	T	I
CREBBP	1109	L	P	F	R	Q	P	V	D	G*	I*	P*	D*	Y*	A*	Y*	N*	V*
P300	1073	L	P	F	R	Q	P	V	D	G*	I*	P*	D*	Y*	A*	Y*	N*	V*
ATAD2	1007	R	V	F	T	K	P	V	D	E	V	P	D	Y	A	Y	N	I*
ATAD2B	981	N	I	F	S	K	P	V	D	E	V	S	D	Y	A	Y	N	I*
BRD9	159	G	F	F	A	F	P	V	T	I	A	P	G	Y	A	Y	N	Y
BRD7	154	A	F	F	S	F	P	V	T	I	A	P	G	Y	A	Y	N	Y
BRPF3	612	H	I	F	A	E	P	V	N	E	V	P	D	Y	C	Y	N	F
BRD1	585	R	I	F	A	Q	P	V	S	E	V	P	D	Y	C	Y	N	F
BRPF1-1	651	N	I	F	S	E	P	V	P	E	V	P	D	Y	C	Y	N	F
BRPF1-2	651	N	I	F	S	E	P	V	P	E*	V*	P*	D*	Y*	C*	Y*	N*	F*
BAZ1A	1452	W	P	F	L	K	L	V	S	Q	V	P	D	Y	C	Y	N	E
CECR2	457	W	P	F	L	E	P	V	D	Y	A	P	N	Y	C	Y	N	Y
BAZ2A	1816	W	P	F	L	E	P	V	N	L	V	S	G	Y	C	F	N	V
BAZ2B	2083	W	P	F	L	L	P	V	N	L	V	P	G	Y	C	F	N	I
TAF1(1)	1403	Y	P	F	H	T	P	V	N	V	V	K	D	Y	S	Y	N	L
TAF1L(1)	1422	H	P	F	H	T	P	V	N	V	V	K	D	Y	S	Y	N	L
TAF1(2)	1526	W	P	F	H	H	P	V	N	F	V	P	D	Y	S	Y	N	Y
TAF1L(2)	1545	W	P	F	H	H	P	V	N	F	V	P	D	Y	S	Y	N	Y
ZMYND8	171	D	A	F	Q	K	P	V	P	Q	H	P	D	Y	C	Y	N	L
PB1(1)	70	E	L	F	I	R	A	P	K	N	Q	P	D	Y	A	Y	Y	E
PB1(2)	206	E	L	F	Q	K	L	P	S	Q	Y	P	D	Y	A	Y	N	V
PB1(3)	406	E	P	F	Y	H	L	P	S	K	Y	P	D	Y	A	Y	N	I
PB1(4)	544	D	L	F	M	V	K	P	S	D	Y	P	D	Y	A	Y	N	V
PB1(5)	682	A	I	F	L	R	L	P	S	E	L	P	D	Y	A	Y	N	I
PB1(6)	798	R	C	Y	S	D	S	L	A	P	A	V	D	P	A	M	N	I
ASH1L	2469	A	P	L	L	N	L	P	S	K	N	A	D	Y	A	Y	Y	V
SMARCA2-1	1425	E	V	F	I	Q	L	P	S	E	L	P	E	Y	A	F	N	I
SMARCA2-2	1407	E	V	F	I	Q	L	P	S	E	L	P	E	Y	A	F	N	I
SMARCA4	1483	E	V	F	I	Q	L	P	S	E	L	P	E	Y	A	F	N	I
Conservation Score		0.33	0.54	0.95	0.35	0.42	0.71	0.71	0.48	0.37	0.42	0.71	0.68	0.95	0.57	0.85	0.86	0.52

The remaining 49 bromodomains were realigned, the conservation scores were determined using Scorecons<sup>220</sup>, and those residues comprising the druggable pocket at the acetyl-lysine binding site were selected for review by consulting KAT2A and KAT2B PDB structures 3D7C<sup>238</sup>



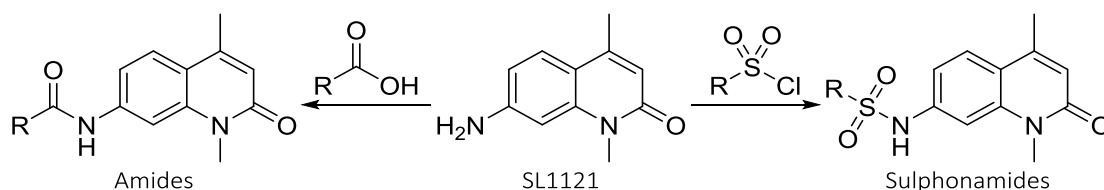
and 5FE8<sup>240</sup>, Table 6.3. Three residues were identified within the ZA loop which are poorly conserved amongst bromodomains and therefore may enable selectivity towards KAT2A and KAT2B, Figure 6.11. The tryptophan, KAT2A W751, situated at the C-terminus of  $\alpha_Z$ , forms part of a lipophilic shelf, termed the WPF (Trp-Pro-Phe) shelf, which is conserved within a number of bromodomains including the type I and BET BRDs. This WPF shelf has previously been targeted to afford potent and selective BET BRD inhibitors.<sup>560</sup> A tool inhibitor designed to interact with this tryptophan should select against many of the bromodomains lacking the WPF shelf, but not the BET BRDs. The methionine, KAT2A M754, exhibits significant diversity throughout the BRDs and therefore could afford good selectivity for KAT2A and KAT2B. Unfortunately, the residue is located a significant distance from the acetyl-lysine binding pocket, which may cause problems with ligand size. Finally, the glutamic acid, KAT2A E761, is poorly conserved throughout the BRDs, including, importantly, the BET family. This residue is located reasonably close to the acetyl-lysine binding pocket and may afford good selectivity for KAT2A and KAT2B.



**Figure 6.11. Poorly Conserved Residues at KAT2A and KAT2B BRD Binding Sites.** KAT2A and KAT2B bromodomains with protein chains coloured from blue at N-terminus to red at C-terminus. Poorly conserved binding site residues highlighted. (a) *Homo sapiens* GCN5 bromodomain (PDB ID: 3D7C)<sup>238</sup>, aligned with that of *S. cerevisiae* (PDB ID: 1E6I)<sup>239</sup>, to show acetyl-lysine binding site. Histone H4 peptide shown in grey. (b) *Homo sapiens* PCAF bromodomain in complex with SL1126, shown in pink (PDB ID: 5FE8).<sup>240</sup>

Comparing PDB structures 3D7C<sup>238</sup> and 5FE8<sup>240</sup>, it was clear that the glutamic acid, KAT2A E761, has the flexibility to move to accommodate small molecules at the binding site, Figure 6.11. This should be advantageous in attempts to grow SL1126 to interact with this residue. Fragment growth at the amide at the 7-position of SL1126 was expected to enable novel interactions with the methionine and glutamic acid residues, and thereby afford selectivity for KAT2A and KAT2B. Fortunately, ligand growth at this secondary amide is theoretically straightforward. Libraries of amides and sulphonamides can be synthesised from SL1121, the primary amide analogue of

SL1126, using carboxylic acids or sulphonyl chlorides respectively, Figure 6.12. This afforded the potential to use automated synthesis to generate compound libraries for screening.

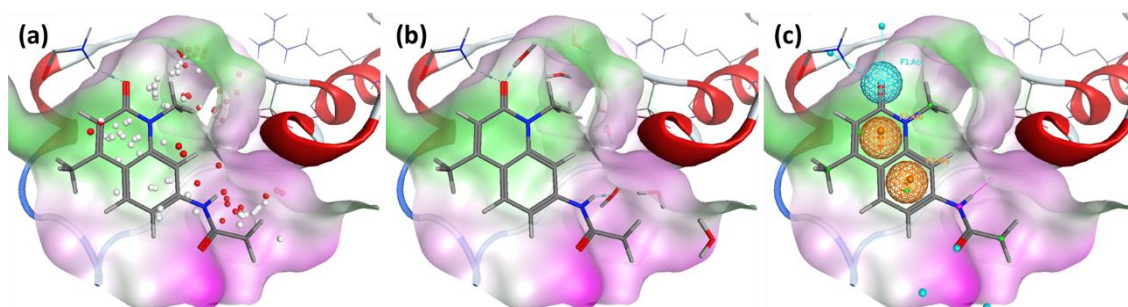


**Figure 6.12. Synthesis of Amide and Sulphonamide Libraries from SL1126.**

Towards the structure-based design of suitable libraries, quinolinone amides and sulphonamides were identified, which could be synthesised from acids available in-house, or acids and sulphonyl chlorides available commercially. A KNIME<sup>242</sup> workflow was designed to generate six conformers of every compound, give each a unique identifier and output an SDF file, which would subsequently be used in docking.

It was decided that the structure of the *Homo sapiens* PCAF BRD in complex with SL1126 (PDB ID: 5FE8)<sup>240</sup> would be utilised in structure-based design. The KAT2A and KAT2B bromodomains are essentially identical and the PCAF structure was preferred as in the apo structure of *Homo sapiens* GCN5 BRD (PDB ID: 3D7C)<sup>238</sup>, K758, which is positioned close to the acetyl-lysine binding pocket, is truncated due to poor resolution. In addition, using the original structure of the complex negated any potential errors induced by docking SL1126 into the GCN5 structure.

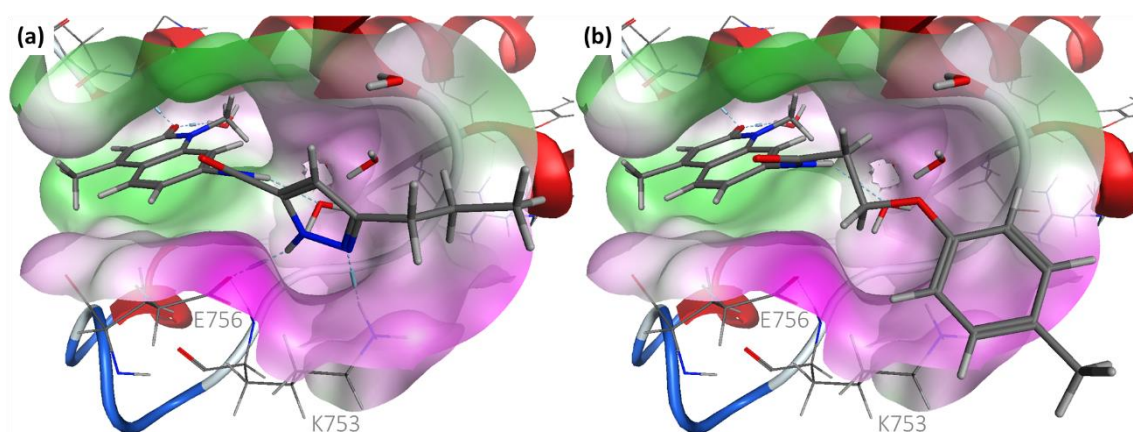
The PCAF BRD (PDB ID: 5FE8)<sup>240</sup> was prepared for docking in MOE.<sup>187</sup> Initially it was protonated, corrected and minimised. Following this, the Site Finder functionality was utilised to identify the site available for ligand binding, dummies were created at the alpha sphere centres, and a molecular surface was generated within 4.5 Å of these dummies, Figure 6.13 (a). A set of conserved waters, located at the active site, were selected that would be incorporated in docking and the remaining solvent was removed, Figure 6.13 (b). Finally, a pharmacophore was generated comprising of the carbonyl acceptor and two aromatic markers at the quinolinone, Figure 6.13 (c).



**Figure 6.13. Preparatory Work for Docking at PCAF BRD.** *Homo sapiens* PCAF bromodomain in complex with SL1126, shown in grey (PDB ID: 5FE8).<sup>240</sup> (a) Ligand binding site depicted by dummy atoms with associated surface coloured by lipophilicity. (b) Conserved water at the binding site depicted as sticks. (c) SL1126 inspired pharmacophore with carbonyl acceptor and two aromatic markers at the quinolinone. Images produced in MOE.<sup>187</sup>

The three libraries of conformers of the amides and sulphonamides were loaded into MOE<sup>187</sup>, prepared, protonated and minimised. The compounds were then docked into the prepared structure of the PCAF BRD using the rigid receptor protocol and employing the pharmacophore as a guide and SL1126 as a template. The results were saved both as MOE molecular database files and SDF files for analysis.

To analyse the results, initially a KNIME<sup>242</sup> workflow was used to group the multiple conformers and associated poses, retaining only the highest scoring pose for each species, sort these by docking score and retrieve the top scoring 40%. These 'top 40%' compounds were analysed by eye in MOE<sup>187</sup> and those which were either interesting because they made new interactions with PCAF, Figure 6.14 (a), or uninteresting because the compound extended completely into the solvent, were marked as such. Following this, another KNIME<sup>242</sup> workflow was designed to split the compounds accordingly. Those compounds that instigated new interactions were combined with the top 20% of compounds that did not instigate new interactions but were reasonable, as they did not protrude completely into the solvent, Figure 6.14 (b), to create the final hit list.



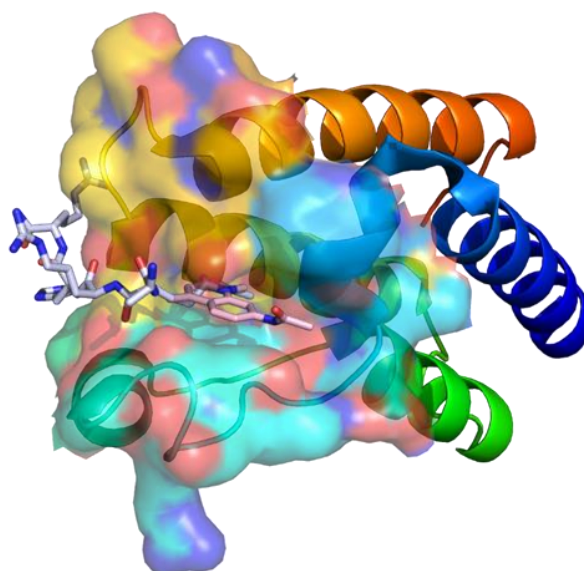
**Figure 6.14. Example Docking Hits.** *Homo sapiens* PCAF bromodomain (PDB ID: 5FE8)<sup>240</sup>, with proposed ligand binding site demonstrated as surface coloured by lipophilicity. (a) Example docking hit exhibiting additional interactions with PCAF BRD. (b) Example docking hit exhibiting no additional electrostatic interactions but also not extending unreasonably into solvent. Images produced in MOE.<sup>187</sup>

From these compound hits, a diverse set were selected for synthesis, comprising of 25 amides from acids available in-house, Figure Apx.8, 100 amides from commercially available acids, Figure Apx.9, and 25 sulphonamides from commercially available sulphonyl chlorides, Figure Apx.10. These libraries were provided to David Pearce (Automated High-Throughput Medicinal Chemistry and Synthesis Lab, University of Sussex) for synthesis.

Unfortunately, synthesis of the amides and sulphonamides proved unexpectedly challenging due to solubility problems. Of the 150 compounds, only 72 were successfully synthesised, 13 amides from acids available in-house, Figure Apx.8, 42 amides from commercially available acids, Figure Apx.9, and 17 sulphonamides from commercially available sulphonyl chlorides,

Figure Apx.10. Disappointingly, a significant number of the compounds that could not be synthesised were those with the lowest molecular weight, which represented the smallest changes from the SL1126 quinolinone starting fragment. These posed less of a challenge to the model than the larger compounds and were most likely to maintain a favourable interaction with the BRD. Regardless, a library of 72 amides and sulphonamides had been successfully designed and synthesised ready for screening against the KAT2A bromodomain.

As an aside, comparing the structure of *Homo sapiens* PCAF BRD (PDB ID: 5FE8)<sup>240</sup> with the *S. cerevisiae* GCN5 BRD in complex with acetylated H4 peptide (PDB ID: 1E6I)<sup>239</sup>, it is evident that there is also space available at the SL1126 4-position, where the acetylated peptide typically interacts with the bromodomain. It is possible that this vector could be exploited, as an alternative to growth at the amide at the 7-position, to improve affinity and specificity, Figure 6.15. No attempts have yet been made to explore SAR at this position and it is believed that this could be valuable.



**Figure 6.15. Alternative Position for SL1126 Fragment Growing.** *Homo sapiens* PCAF bromodomain with protein chain coloured from blue at N-terminus to red at C-terminus in complex with SL1126 in pink (PDB ID: 5FE8)<sup>240</sup>, aligned to *S. cerevisiae* bromodomain (PDB ID: 1E6I)<sup>239</sup> to show acetyl-lysine binding site. Histone H4 peptide in grey.

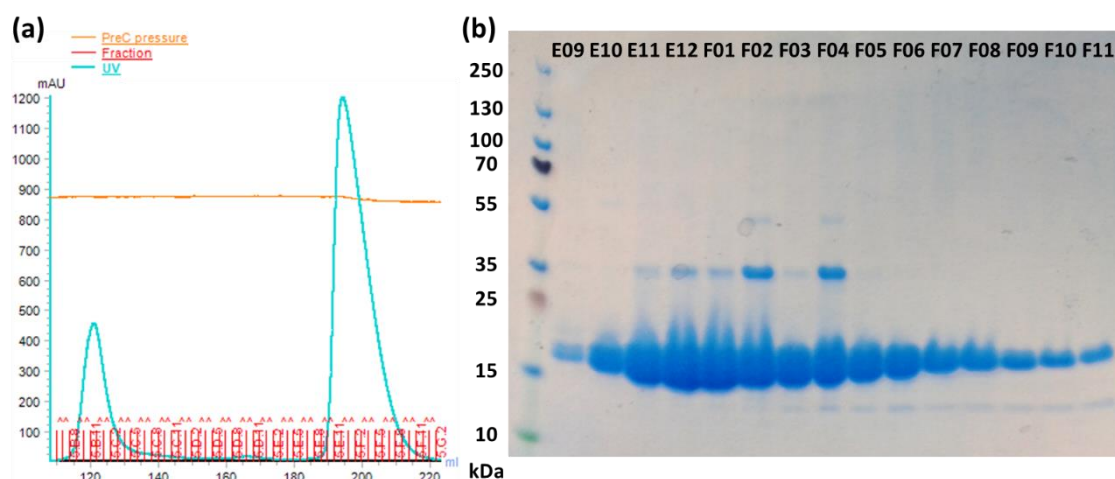
#### 6.2.4 *T<sub>m</sub>* Shift Optimisation and Screening

With a library of potential KAT2A and KAT2B BRD inhibitors ready for testing, a suitable assay was required in-house to begin screening efforts. The thermal stability shift assay was prioritised due to relative ease of assay set up and capacity for high throughput. Thermal stability shift is intended to identify ligands that bind and stabilise proteins. The thermal denaturation temperature at which a protein unfolds is characterised by an increase in fluorescence of an applied dye (SYPRO orange) with affinity for hydrophobic regions that are exposed only upon protein unfolding. An increase in melting temperature (*T<sub>m</sub>*) upon addition of a ligand is



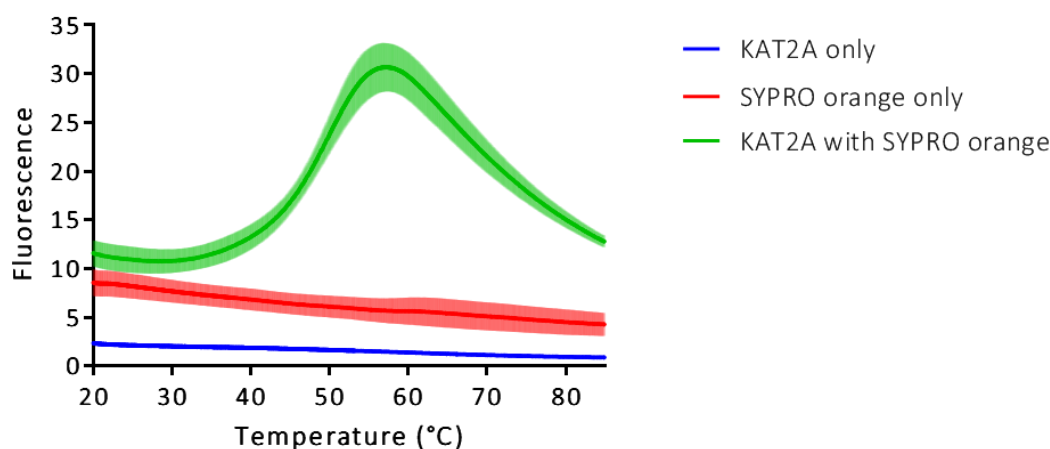
interpreted as a binding event, as the small molecule has increased the energy required to disrupt the protein structure, indicative of a constructive interaction.

Histidine-tagged KAT2A bromodomain protein was required for use in thermal shift screening. The 16 kDa protein was successfully expressed and purified, using immobilised metal affinity chromatography and size exclusion chromatography, from a construct generously provided by Dr Antony Oliver (Genome Damage and Stability Centre, University of Sussex), Figure 6.16.

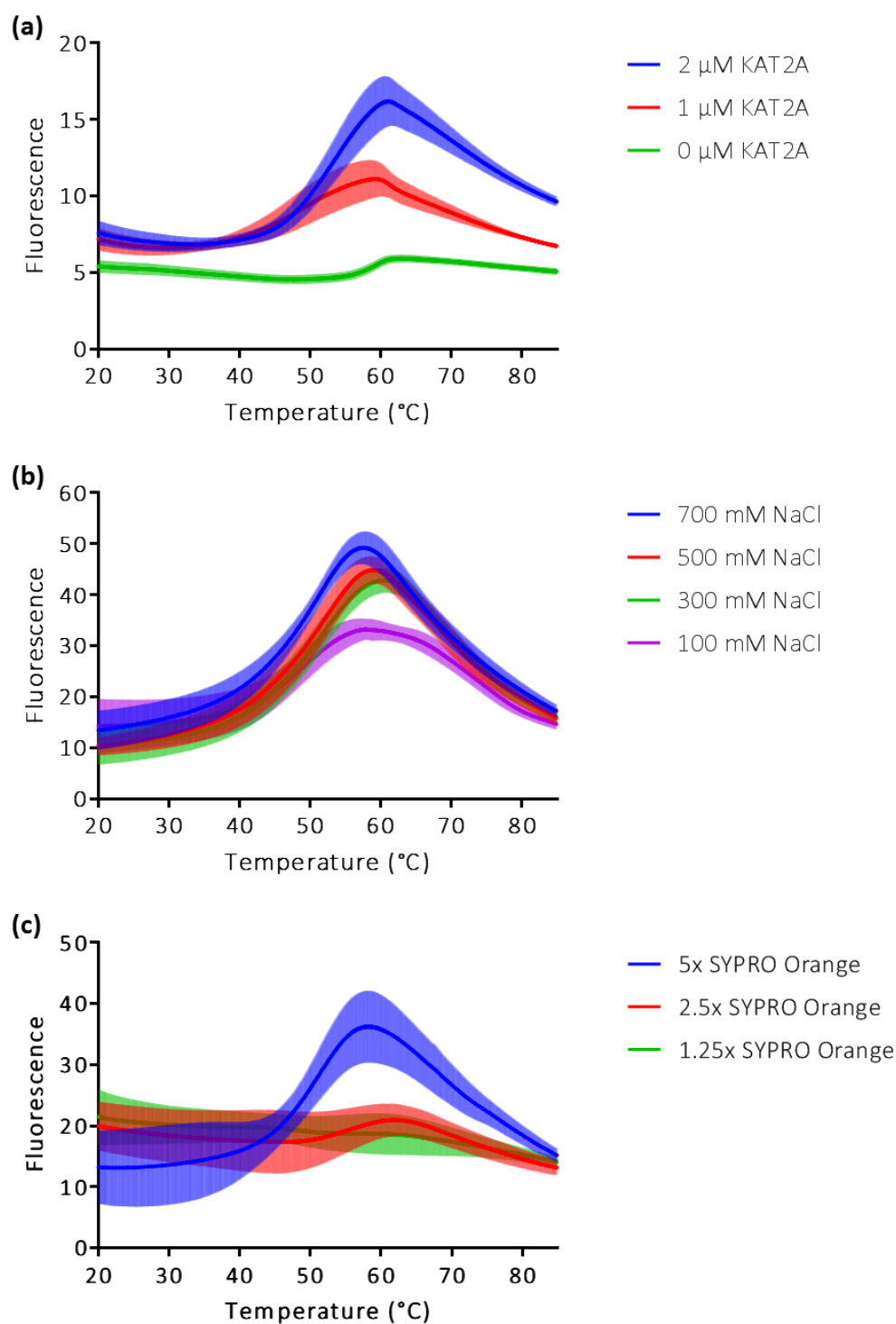


**Figure 6.16. His-Tagged KAT2A BRD Purification.** (a) Chromatogram from SEC purification of His6-KAT2A BRD. (b) SDS-PAGE gel of fractions obtained by SEC, numbered to correspond to the chromatogram. Concentrated protein band at 16 kDa corresponds to His6-KAT2A BRD protein.

After preparation of the His6-KAT2A BRD protein, efforts in  $T_m$  shift assay optimisation were initiated. To begin with, it was necessary to confirm that the thermal denaturation of the KAT2A BRD protein could be successfully measured, Figure 6.17. 2  $\mu$ M KAT2A BRD was added to 5x SYPRO orange dye and the fluorescence was measured whilst ramping the temperature between 20  $^{\circ}$ C and 85  $^{\circ}$ C. A melting temperature of  $\sim$ 50  $^{\circ}$ C was observed.



**Figure 6.17. Thermal Denaturation of KAT2A BRD.** Fluorescence of 5x SYPRO orange dye with or without 2  $\mu$ M KAT2A, 20  $\mu$ L well volume, measured over temperature ramp 20–85  $^{\circ}$ C. Graph represents mean  $\pm$  1 SD of three experiments.

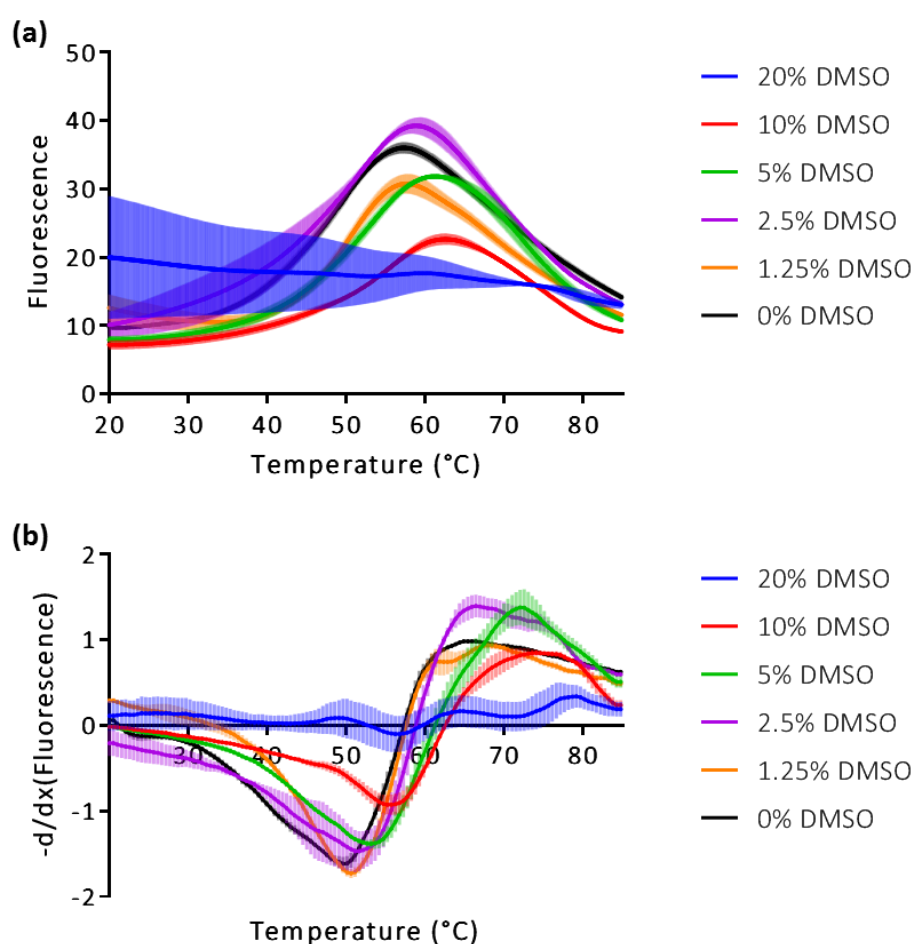


**Figure 6.18. KAT2A BRD Thermal Stability Shift Assay Optimisation.** Fluorescence of SYPRO orange dye with KAT2A and 2% DMSO, 20  $\mu$ L well volume, measured over temperature ramp 20-85 °C. (a) 5x SYPRO orange dye with 0  $\mu$ M, 1  $\mu$ M or 2  $\mu$ M KAT2A, to measure effect of protein concentration. (b) 5x SYPRO orange dye with 2  $\mu$ M KAT2A in buffer with varied concentrations of sodium chloride, to measure effect of salt concentration. (c) 1.25x, 2.5x or 5x SYPRO orange dye with 2  $\mu$ M KAT2A, to measure effect of SYPRO orange concentration. Graphs represent mean  $\pm$  1 SD of three experiments.

Following this, a number of optimisation experiments were conducted. Firstly, the effect of protein concentration was assessed. 0  $\mu$ M, 1  $\mu$ M or 2  $\mu$ M KAT2A BRD was added to 5x SYPRO orange dye, and the fluorescence was measured whilst increasing the temperature from 20 °C

to 85 °C, Figure 6.18 (a). Whilst significant, the fluorescence at 1  $\mu$ M KAT2A was considerably less than the signal at 2  $\mu$ M. Given the ease of protein preparation for this target, it was decided that 2  $\mu$ M KAT2A should be used in the assay as standard. In the absence of protein, minimal fluorescence was observed and there was no significant change with temperature.

Secondly, the effect of buffer composition was assessed. The thermal denaturation of 2  $\mu$ M KAT2A BRD protein was measured in a buffer containing 100 mM, 300 mM, 500 mM or 700 mM sodium chloride, Figure 6.18 (b). The KAT2A BRD showed a stable melting curve at NaCl concentrations of 300 mM or above, but showed destabilisation at 100 mM. It was decided that 500 mM salt should be used as standard, consistent with a protocol provided by the SGC.

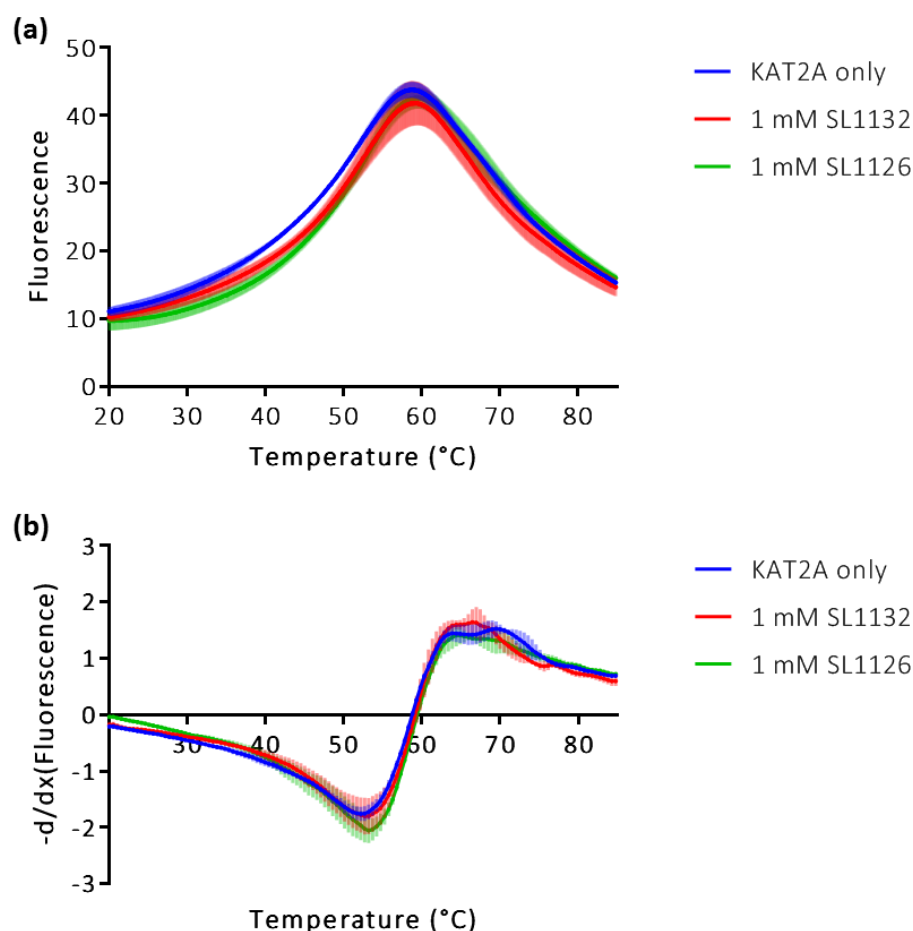


**Figure 6.19. DMSO Tolerance of Thermal Stability Shift Assay.** Fluorescence of 5x SYPRO orange dye with 2  $\mu$ M KAT2A and 0-20% DMSO, 20  $\mu$ L well volume, measured over temperature ramp 20-85 °C. (a) Melting curves. (b) Melting peaks, calculated as the derivative of the melting curve. Minima correspond to T<sub>m</sub>. Graphs represent mean  $\pm$  1 SD of three experiments.

Thirdly, the effect of the concentration of the SYPRO orange dye was assessed. 2  $\mu$ M KAT2A BRD was added to 1.25x, 2.5x or 5x SYPRO orange dye and the fluorescence was measured whilst increasing the temperature between 20 °C and 85 °C, Figure 6.18 (c). Decreasing the concentration of SYPRO orange significantly reduced the fluorescence signal and therefore the

sensitivity of the assay. It was determined that the dye should be used at 5x concentration as standard, again consistent with the protocol provided by the SGC.

Finally, the impact of addition of DMSO was assessed. The small molecules tested in assays are rarely water soluble and therefore must be dissolved in a suitable solvent prior to addition. DMSO is popular as it able to dissolve both polar and nonpolar compounds and is miscible with many organic solvents as well as water. Consequently, most compound libraries, including all those available in-house are dissolved in DMSO, so it is necessary to test the DMSO tolerance of protein assays. Thermal denaturation of the KAT2A BRD protein was measured in the presence of 0-20% DMSO, Figure 6.19 (a). Stable melting curves were observed for the KAT2A BRD with up to 5% DMSO. At 20% DMSO the protein appeared to be completely denatured. Notably, it was also observed that addition of any concentration of DMSO induced a shift in melting temperature, e.g. from 50 °C at 0% DMSO to 53 °C at 5%, Figure 6.19 (b). This was expected, as DMSO is a known BRD ligand and will thereby affect protein stability.<sup>558</sup> Accordingly, any comparisons in  $T_m$  should be made only between samples of equivalent DMSO concentration.

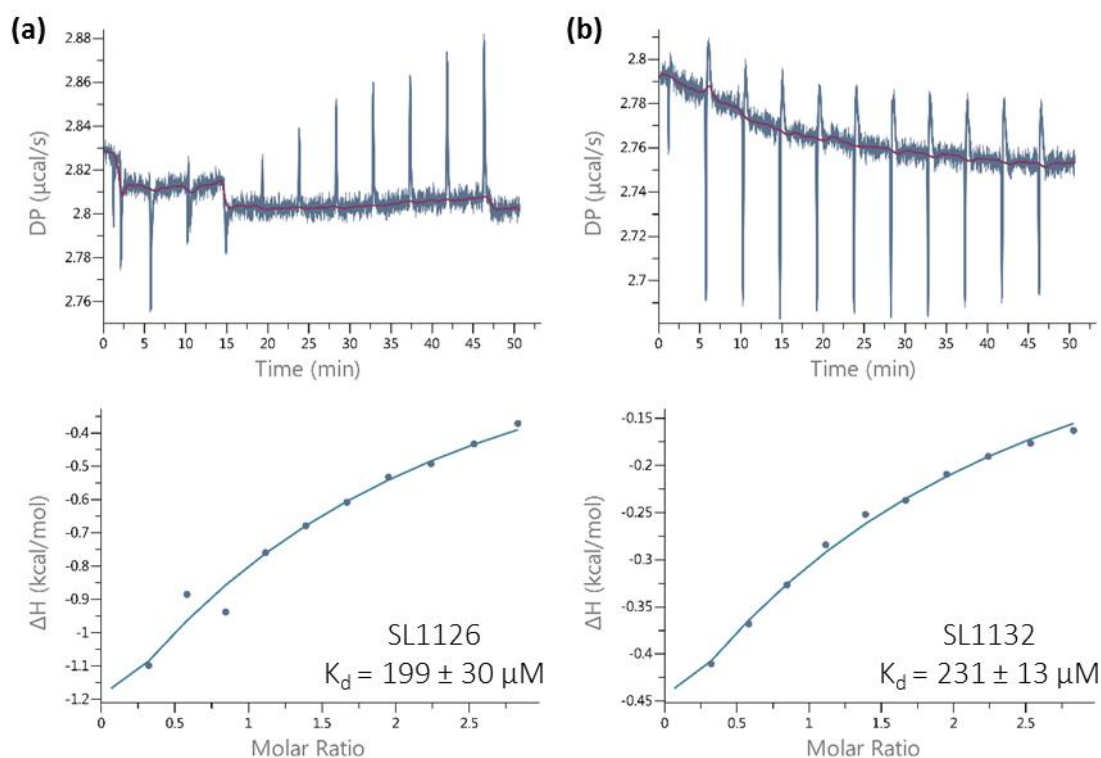


**Figure 6.20. KAT2A BRD  $T_m$  Shift with SL1126 and SL1132.** Fluorescence of 5x SYPRO orange dye with 2  $\mu$ M KAT2A and 2% DMSO, with or without 1 mM SL1126 or SL1132, 20  $\mu$ L well volume, measured over temperature ramp 20-85 °C. (a) Melting curves. (b) Melting peaks, calculated as the derivative of the melting curve. Minima correspond to  $T_m$ . Graphs represent mean  $\pm$  1 SD of three experiments.



Having optimised the thermal shift assay conditions, an attempt was made to reproduce some of the results provided by the SGC. The thermal denaturation of 2  $\mu\text{M}$  KAT2A BRD protein was measured in the presence and absence of 1 mM SL1126 or SL1132 with 2% DMSO, Figure 6.20, and the melting temperatures, determined using the Roche Protein Melting Analysis Software, were compared to calculate the  $T_m$  shift. Frustratingly, the  $T_m$  shifts were measured as just 1.0  $^{\circ}\text{C}$  and 0.5  $^{\circ}\text{C}$  for SL1126 and SL1132 respectively, considerably lower than the 5.7  $^{\circ}\text{C}$  and 4.2  $^{\circ}\text{C}$  shifts observed of KAT2B by the SGC.

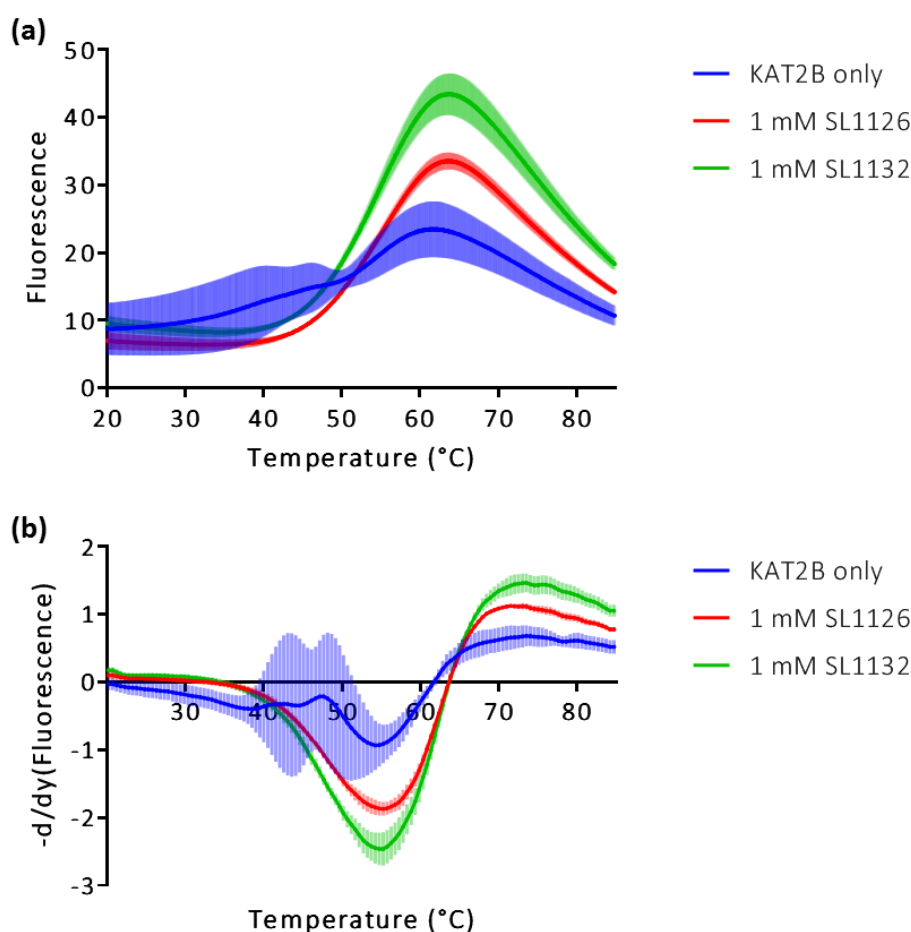
To investigate the cause of this discrepancy, isothermal titration calorimetry was employed to measure the heats of binding of SL1126 and SL1132 with the KAT2A BRD. The results of experiments conducted at 800  $\mu\text{M}$  compound and 60  $\mu\text{M}$  protein are shown in Figure 6.21. SL1126 exhibited a  $K_d$  of 199  $\mu\text{M}$  and SL1132 a  $K_d$  of 231  $\mu\text{M}$ , Table 6.4. These values were again considerably higher than those reported with KAT2B by the SGC, 30  $\mu\text{M}$  and 35  $\mu\text{M}$  respectively.



**Figure 6.21. ITC Binding Data for Interaction Between KAT2A BRD and SL1126 or SL1132.** Isotherms of raw titration heat above. Normalised binding heats below, with solid line representing nonlinear least squares fit using single-site binding model. Error calculated from deviation to least squares fit. (a) 800  $\mu\text{M}$  SL1126 titrated into 60  $\mu\text{M}$  KAT2A BRD. (b) 800  $\mu\text{M}$  SL1132 titrated into 60  $\mu\text{M}$  KAT2A BRD.

Based on the thermal shift and ITC data it appeared that SL1126 and SL1132 were significantly less potent inhibitors of the KAT2A BRD than expected. It was rationalised that this observation could be explained in two ways. Either the compounds were less potent inhibitors of the KAT2B bromodomain than reported, or the compounds exhibited a considerable, unexpected selectivity for KAT2B over KAT2A. To test this, the thermal denaturation of 2  $\mu\text{M}$  KAT2B BRD

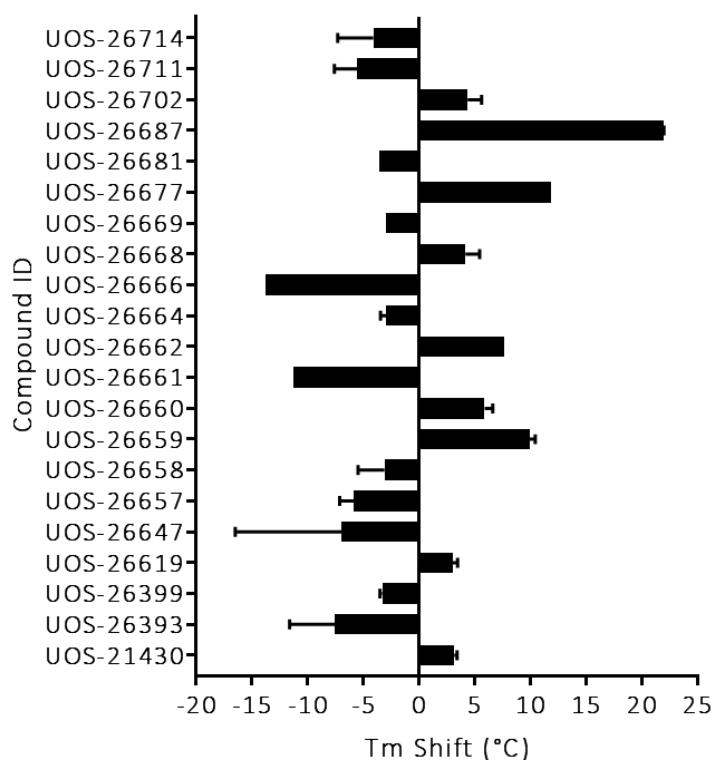
protein, generously provided by the SGC, was measured in the presence and absence of 1 mM SL1126 or SL1132 with 2% DMSO, Figure 6.22, and the melting temperatures, determined using the Roche Protein Melting Analysis Software, were compared to calculate the  $T_m$  shift. The  $T_m$  shifts were measured as 2.5 °C and 2.1 °C for SL1126 and SL1132 respectively, higher than observed for the KAT2A BRD but still lower than originally reported by the SGC. It was therefore concluded that the inhibitors might exhibit some selectivity for KAT2B compared to KAT2A, but fundamentally, they were not as potent as previously believed. Correspondingly, when it was ultimately published that SL1126 inhibits PCAF, a thermal shift of 2.6 °C was reported and no ITC data was included.<sup>240</sup>



**Figure 6.22. KAT2B BRD  $T_m$  Shift with SL1126 and SL1132.** Fluorescence of 5x SYPRO orange dye with 2  $\mu$ M KAT2B and 2% DMSO, with or without 1 mM SL1126 or SL1132, 20  $\mu$ L well volume, measured over temperature ramp 20–85 °C. (a) Melting curves. (b) Melting peaks, calculated as the derivative of the melting curve. Minima correspond to  $T_m$ . Graphs represent mean  $\pm$  1 SD of three experiments.

This result served to bolster confidence that the thermal stability shift assay had been appropriately optimised for use in screening. However, SL1126 no longer represented such a desirable lead compound for structure-based design, which boded poorly for the libraries of amides and sulphonamides inspired by this scaffold.

Despite these concerns, the libraries of amides and sulphonamides, along with a handful of other quinolinone homologues generously prepared by Kamlesh Bala (Sussex Drug Discovery Centre, University of Sussex), were screened against the KAT2A BRD using the optimised T<sub>m</sub> shift assay. The thermal denaturation of 2  $\mu$ M KAT2A BRD was measured in the presence and absence of 1 mM inhibitor with 2% DMSO, and the melting temperatures, determined using the Roche Protein Melting Analysis Software, were compared to calculate the T<sub>m</sub> shifts. Those compounds which induced a significant shift in T<sub>m</sub>, calculated as greater than three times the standard deviation in the controls, which equated to approximately 2.5  $^{\circ}$ C, are displayed in Figure 6.23. It is notable that within these constraints, the binding of SL1126 and SL1132 would not have been detected. This is illustrative of the poor sensitivity of this thermal shift assay.

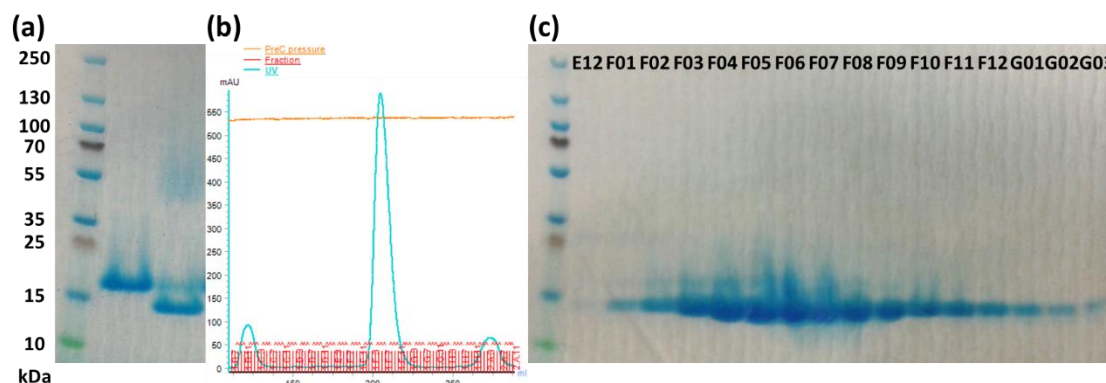


**Figure 6.23. KAT2A BRD T<sub>m</sub> Shift Screening Results.** T<sub>m</sub> shift values calculated by comparing T<sub>m</sub> of protein with and without compound, as determined by Roche Protein Melting Analysis Software. Fluorescence of 5x SYPRO orange dye with 2  $\mu$ M KAT2A, 2% DMSO and 1 mM compound, 20  $\mu$ L well volume, measured over temperature ramp 20-85  $^{\circ}$ C. Graph represents mean  $\pm$  1 SD of two experiments.

Nine compounds induced a significant positive T<sub>m</sub> shift and were of interest. The positive shift indicated an increase in the stability of the KAT2A BRD upon compound addition, which it was hoped corresponded to protein binding. Twelve compounds exhibited a significant negative T<sub>m</sub> shift, apparently decreasing the stability of the protein. These compounds may interact favourably with the denatured protein and thereby shift the equilibrium toward the unfolded state. These were not of interest in this work.

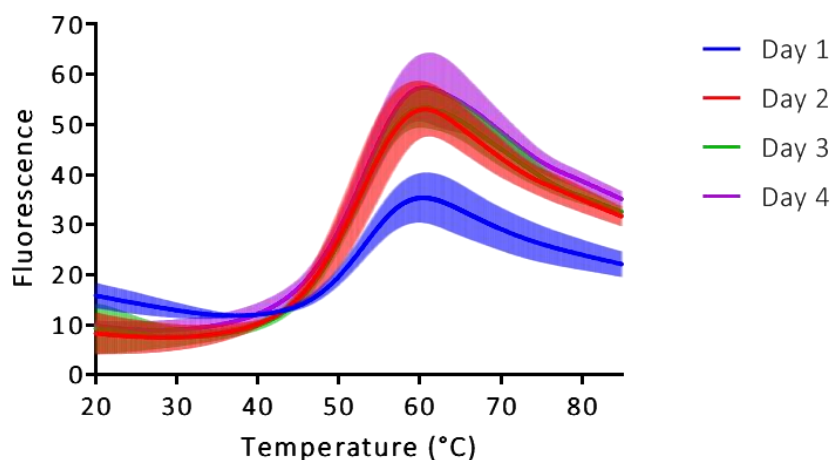
### 6.2.5 NMR Preparation and Screening

Whilst growing and diversifying the SL1126 fragment would enable exploration of SAR, as previously discussed, the quinolinone fragments show no selectivity for the type I BRDs over other bromodomain families. As such, development of a selective small molecule inhibitor of the KAT2A bromodomain required screening to identify other suitable cores. NMR fragment screening was selected as a suitable technique for identification of novel scaffolds for inhibitor development.



**Figure 6.24. KAT2A BRD Purification.** (a) SDS-PAGE gel showing His6-KAT2A BRD (16 kDa) and KAT2A BRD (13.5 kDa) indicating successful cleavage by TEV protease. (b) Chromatogram from SEC purification of KAT2A BRD in sodium phosphate buffer. (c) SDS-PAGE gel of fractions obtained by SEC, numbered to correspond to the chromatogram. Concentrated protein band at 13.5 kDa corresponds to KAT2A BRD protein.

The NMR fragment screen was conducted externally but some preparatory work was conducted in-house. The assay required untagged KAT2A bromodomain protein, purified in a sodium phosphate buffer. Accordingly, the His6-KAT2A bromodomain protein was expressed and purified and the histidine tag was cleaved using TEV (tobacco etch virus) protease to afford the 13.5 kDa KAT2A BRD protein, Figure 6.24.



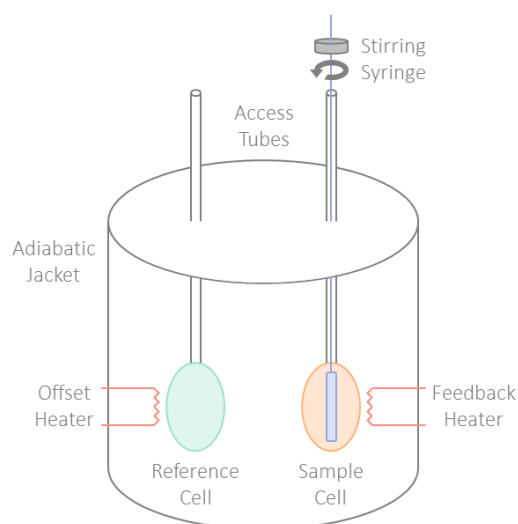
**Figure 6.25. KAT2A BRD Protein Stability Test.** Thermal denaturation of KAT2A BRD stored at 4 °C in 20 mM sodium phosphate buffer with 200 mM NaCl, measured over four days. Fluorescence of 5x SYPRO orange dye with 2 μM KAT2A BRD, 20 μL well volume, measured over temperature ramp 20-85 °C. Graph represents mean  $\pm$  1 SD of triplicate repeats.

It was necessary to confirm that the protein would be stable at 4 °C for the duration of the NMR assay, which runs over four days, ideally in a buffer with lower than the standard 500 mM sodium chloride. A solution of 2  $\mu$ M KAT2A BRD in sodium phosphate buffer with 200 mM NaCl was stored for four days in the fridge. Every 24 hours a sample of the protein was taken and the thermal denaturation was measured. The protein was stable over this period, Figure 6.25.

The NMR fragment screening was conducted externally utilising Water-Ligand Observed via Gradient Spectroscopy (WaterLOGSY) and Saturation Transfer Difference (STD) NMR techniques. 32 fragments were identified that exhibited affinity for the KAT2A BRD, Figure Apx.11. These compounds were purchased for confirmation in-house.

#### 6.2.6 ITC Hit Confirmation

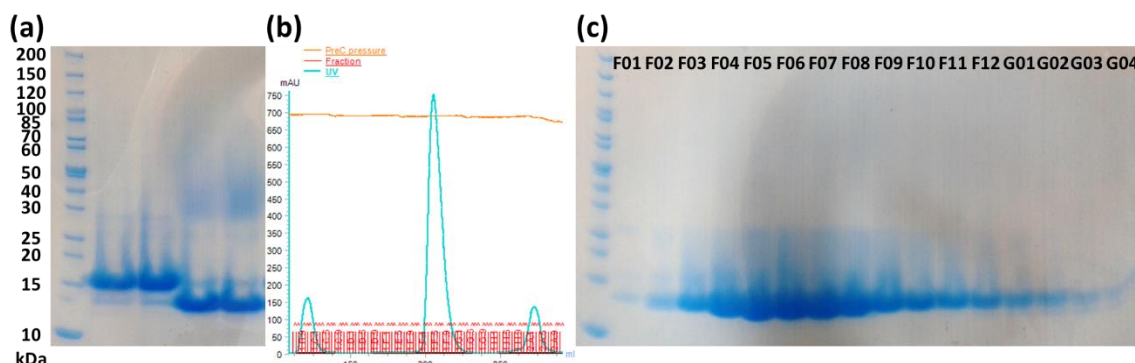
Hit confirmation was conducted using isothermal titration calorimetry. ITC is a powerful and sensitive technique for characterisation of the thermodynamics of molecular interactions in solution. Typically, a test compound is titrated, via a series of small injections, into the protein sample. Any interaction between the compound and the protein will cause a small release or uptake of heat within the sample cell, which is measured relative to a reference cell, Figure 6.26. Monitoring of the relationship between the dose of the compound added and the magnitude of the heat changes enables calculation of the equilibrium dissociation constant ( $K_d$ ), which corresponds to the free ligand concentration when 50% of the protein binding sites are occupied.<sup>572</sup> This ability to directly monitor binding means that ITC can be used to characterise protein fragments which lack native biochemical activity, such as bromodomains.



**Figure 6.26. Schematic of Isothermal Titration Calorimetry Instrumentation.**

Consulting with the SGC it was discovered that untagged BRD protein is utilised as standard in ITC experiments, with the histidine tag cleaved from the protein during purification. Reportedly,

the tag can be problematic at the high protein concentrations required in ITC, impacting protein stability and causing aggregation. Accordingly, KAT2A bromodomain protein was expressed and purified with the histidine tag cleaved, affording the 13.5 kDa protein, Figure 6.27.



**Figure 6.27. KAT2A BRD Purification.** (a) SDS-PAGE gel showing His6-KAT2A BRD (16 kDa) and KAT2A BRD (13.5 kDa) indicating successful cleavage by TEV protease. (b) Chromatogram from SEC purification of KAT2A BRD. (c) SDS-PAGE gel of fractions obtained by SEC, numbered to correspond to the chromatogram. Concentrated protein band at 13.5 kDa corresponds to KAT2A BRD protein.

ITC requires no specialised reagents or labelling of reactants and the same detection method is used for all protein targets. As a result, assay development is minimal. In addition, interpretation of measured binding affinities is straightforward as there are no complicating effects, for example from substrates, intermediates, products or reporter ligands.<sup>572</sup>

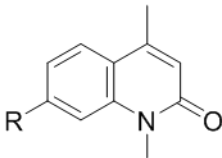
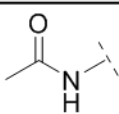
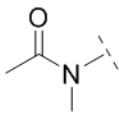
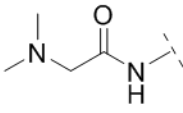
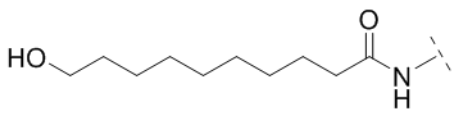
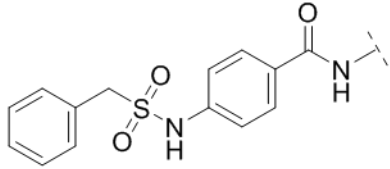
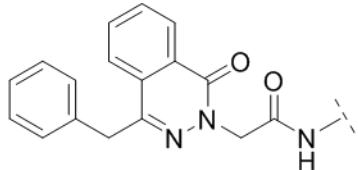
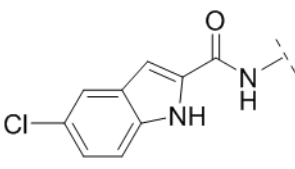
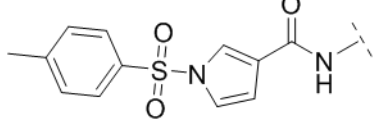
Initial attempts to confirm the affinity of the amides identified in the thermal shift screen were hindered by solubility difficulties. The compounds were not soluble at the necessary concentration ( $\geq 600 \mu\text{M}$ ) to enable injection into the test protein. Fortunately, however, it was discovered that the KAT2A BRD protein was soluble at this concentration. It was therefore determined that the compound-protein addition should be reversed, such that KAT2A BRD was injected into a solution of the small molecule. Fortunately, this also aided in improving the throughput of the assay. A single heat of dilution, of protein into buffer, could be conducted as the control experiment for titration of that sample of KAT2A BRD into a series of compounds.

Unfortunately, UOS-26687, UOS-26702 and UOS-26619 remained insoluble at  $60 \mu\text{M}$  and therefore could not be tested. However, the heats of binding of the other amides with the KAT2A BRD were measured. Only compounds UOS-21430 and UOS-26660 exhibited binding with the target protein, Table 6.4. UOS-21430 exhibited a  $K_d$  of  $29 \mu\text{M}$  and UOS-26660 a  $K_d$  of  $170 \mu\text{M}$ . The results of the experiments conducted at  $613 \mu\text{M}$  KAT2A BRD and  $60 \mu\text{M}$  compound are shown in Figure 6.28. UOS-21430 represented an interesting hit as it represented only a minor change from SL1126, with addition of a methyl, and therefore retained good ligand efficiency.

Overall, this was a disappointing result, exposing a high frequency of false positives from the  $T_m$  shift screen. Of the compounds that exhibited a positive  $T_m$  shift and could be tested by ITC,

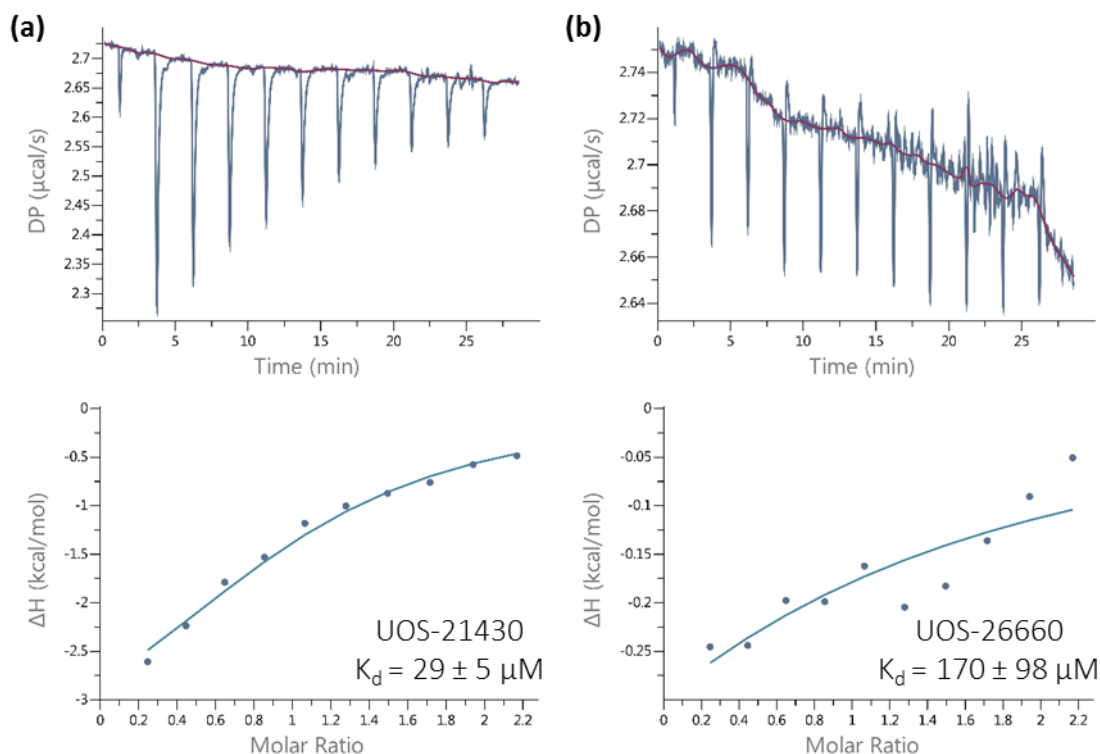
only 33% confirmed as active. With the poor sensitivity and high rate of false positives, it was concluded that the T<sub>m</sub> shift assay did not represent a useful screening technique for KAT2A BRD.

**Table 6.4. T<sub>m</sub> Shift and ITC K<sub>d</sub> Data From Experiments with KAT2A BRD and Amides Designed as KAT2A/KAT2B Binders.** Error for ITC data calculated from deviation to least squares fit.

			
Compound	R	T <sub>m</sub> Shift at 1 mM (°C)	ITC K <sub>d</sub> (μM)
SL1126		1	199 ± 30
UOS-21430		3.2	29 ± 5
SL1132		0.5	231 ± 13
UOS-26659		10.0	> 500
UOS-26660		5.9	170 ± 98
UOS-26662		7.7	> 500
UOS-26668		4.2	> 500
UOS-26677		11.9	> 500

Additionally, the result represented a low hit rate from the libraries of compounds designed from the SL1126 lead. This was somewhat anticipated when the true affinity of SL1126 for the KAT2A BRD was determined. Furthermore, the poor solubility associated with the quinolinones was highly detrimental to fragment growth efforts and ultimately defined which compounds

could be tested effectively. It was critically important that a novel scaffold be identified, which offered improved scope for growth and optimisation.



**Figure 6.28. ITC Binding Data for Interaction Between KAT2A BRD and UOS-21430 or UOS-26660.** Isotherms of raw titration heat above. Normalised binding heats below, with solid line representing nonlinear least squares fit using single-site binding model. Error calculated from deviation to least squares fit. (a) 613  $\mu\text{M}$  KAT2A BRD titrated into 60  $\mu\text{M}$  UOS-21430. (b) 613  $\mu\text{M}$  KAT2A BRD titrated into 60  $\mu\text{M}$  UOS-26660.

Towards this aim, ITC was used as a secondary assay to confirm the hits from the NMR fragment screen. Fragment libraries consist of small molecules of molecular weight lower than 300 Da. Because of this small size, fragment space is smaller than chemical space and therefore can be more effectively probed using a smaller library. The drawback of this low molecular weight is that fragments are typically less potent than larger compounds, thus requiring highly sensitive assay techniques to detect binding.

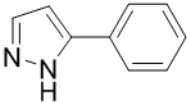
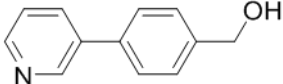
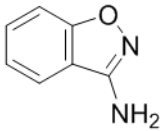
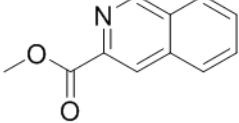
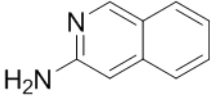
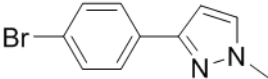
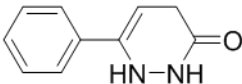
The heats of KAT2A BRD binding of the 32 fragments identified by NMR were measured, again by titrating protein into a solution of the small molecule, to maximise throughput. Seven of the fragments exhibited binding, Table 6.5. The results of the experiments conducted at 600  $\mu\text{M}$  KAT2A BRD with 60  $\mu\text{M}$  UOS-18715 and UOS-28671 are shown in Figure 6.29.

Although dissociation constants were calculated, these were associated with high error, and there is a lack of confidence in the values. The low affinity of the fragments had two effects. Firstly, the measured heats of binding were small. As well as increasing the error due to background noise, this permitted the small perturbations resulting from binding of DMSO, a known BRD ligand<sup>558</sup>, to impact the data. This results in a poor fit between the experimental



data and the nonlinear least squares fit using a single-site binding model. Secondly, the titrations did not represent full binding curves. In a full curve, at low molar ratios, 100% of the ligand will bind to the available protein and eventually, when the protein becomes saturated, 0% of the ligand will bind. Commonly, for compounds of low affinity, only a section of this binding curve is observed, for example from 90% binding to 10% binding, and therefore the data must be extrapolated to calculate the dissociation constant, inducing error in the result. It is likely that in the case of these fragments the titrations represent only a small section of the binding curve, affording significant error. When a ligand with low micromolar affinity, for which a full binding curve can be obtained, becomes available, competition assays could be employed to more accurately calculate  $K_d$ s for these weakly binding ligands.

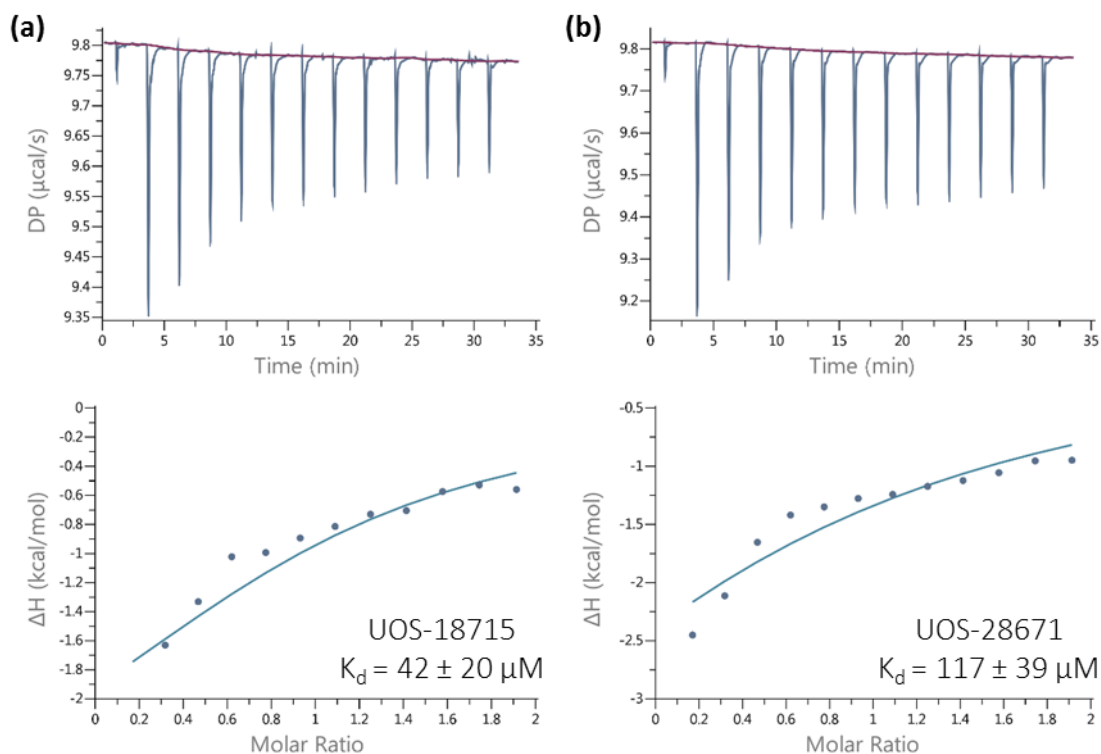
**Table 6.5.  $K_d$  Data from ITC Experiments with KAT2A BRD and Fragments Identified as KAT2A Binders by NMR.** Error calculated from deviation to least squares fit.

Compound	Structure	ITC $K_d$ ( $\mu$ M)
UOS-10321		$111 \pm 54$
UOS-10879		$27 \pm 23$
UOS-11402		$50 \pm 45$
UOS-18715		$42 \pm 20$
UOS-28671		$117 \pm 39$
UOS-28675		$12 \pm 11$
UOS-28676		$13 \pm 11$

### 6.2.7 ITC Screening

Having concluded that the thermal stability shift assay exhibited prohibitively poor sensitivity and high rate of false positives, ITC was instead adopted as the primary screening assay. The principal limitation of ITC is the high quantity of test protein required.<sup>572</sup> Fortunately, preparation of the KAT2A BRD was straightforward and sufficient protein could be generated.

ITC is also not a high-throughput technique, with each titration taking at least 30 minutes. To maximise throughput, KAT2A BRD protein was titrated into compound as standard.



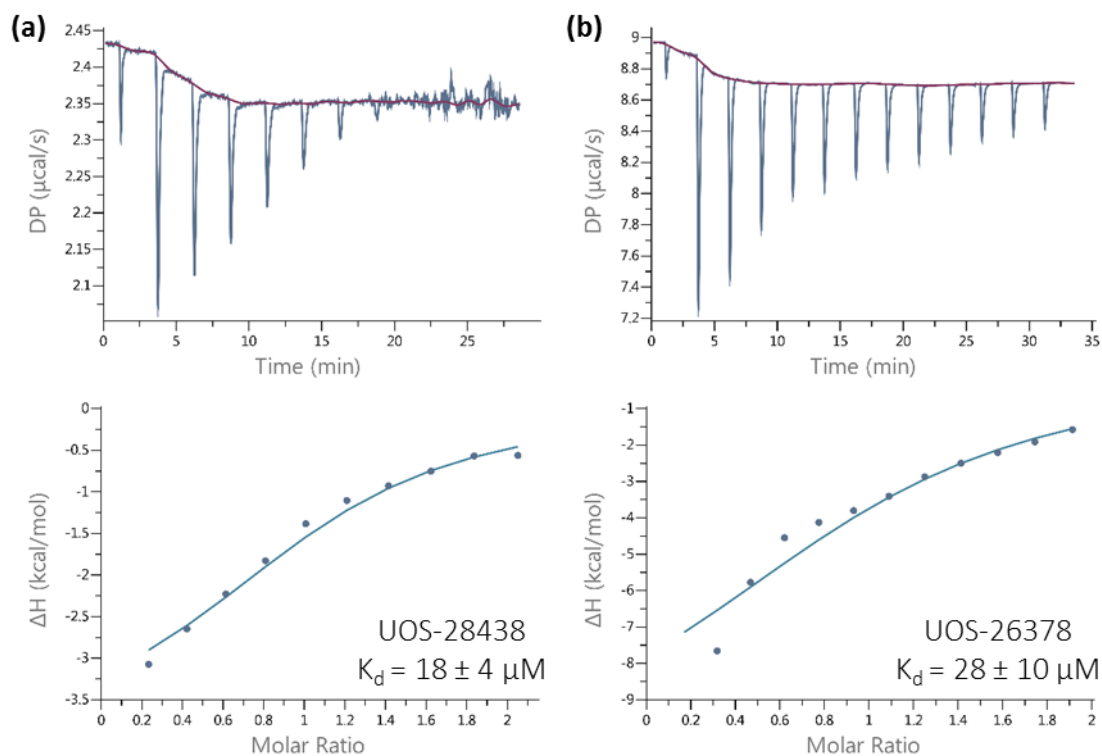
**Figure 6.29.** ITC Binding Data for Interaction Between KAT2A BRD and UOS-18715 or UOS-28671. Isotherms of raw titration heat above. Normalised binding heats below, with solid line representing nonlinear least squares fit using single-site binding model. Error calculated from deviation to least squares fit. (a) 600  $\mu\text{M}$  KAT2A BRD titrated into 60  $\mu\text{M}$  UOS-18715. (b) 600  $\mu\text{M}$  KAT2A BRD titrated into 60  $\mu\text{M}$  UOS-28671.

A series of quinolinones inspired by SL1126 and UOS-21430, prepared by Dr Darren Le Grand and Kamlesh Bala (Sussex Drug Discovery Centre, University of Sussex) were screened by ITC. Seven of the compounds exhibited binding, Table 6.6. Apart from UOS-28438, which exhibited a  $K_d$  of 18  $\mu\text{M}$ , none of the compounds showed a significant increase in potency compared to UOS-21430 ( $K_d = 29 \mu\text{M}$ ). The results of the experiments with  $\sim 600 \mu\text{M}$  KAT2A BRD and 60  $\mu\text{M}$  of the two most potent compounds, UOS-28438 and UOS-26378, are shown in Figure 6.30. This data fits poorly with the binding model, as was observed with fragments UOS-18715 and UOS-28671. This is likely a result of DMSO binding effects.

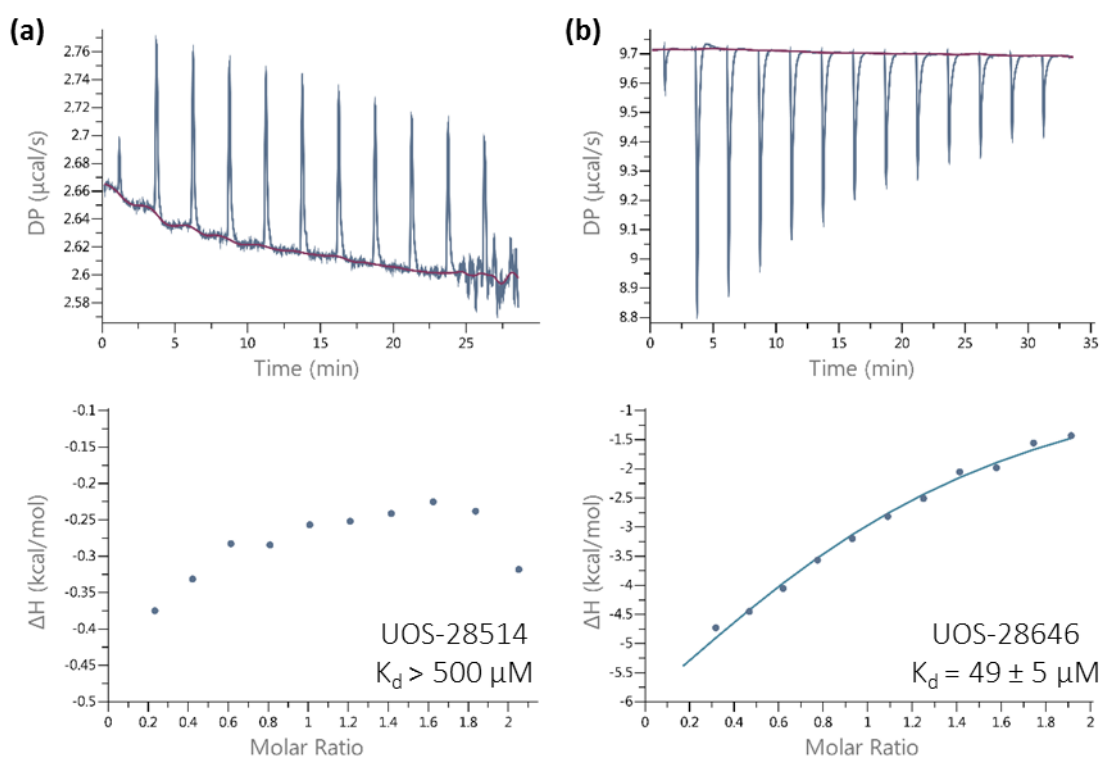
UOS-28438 was an interesting hit. Considering the crystal structure of SL1126 in complex with PCAF (PDB ID: 5FE8)<sup>240</sup>, the ethyl group at the 1-position is expected to extend towards the conserved waters at the back of the binding cavity, Figure 6.13 (b). There are limited examples in the literature in which any of these waters are displaced.<sup>573</sup> It was therefore surprising that binding was retained with ligand growth at this position. It was speculated that the binding position of the ligand may be adjusted to accommodate the ethyl group. UOS-28438 was therefore highlighted as an interesting target for crystallographic study.

**Table 6.6.  $K_d$  Data from ITC Experiments with KAT2A BRD and Quinolinone Derivatives.** Error calculated from deviation to least squares fit.

Compound	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	ITC $K_d$ ( $\mu$ M)
UOS-26391	CH <sub>3</sub>	CH <sub>3</sub>	H	> 500
UOS-28450	CH <sub>3</sub>	CH <sub>3</sub>	Br	34 $\pm$ 3
UOS-26378	CH <sub>3</sub>	CH <sub>3</sub>		28 $\pm$ 10
UOS-28000	CH <sub>3</sub>	CH <sub>3</sub>		71 $\pm$ 7
UOS-28001	CH <sub>3</sub>	CH <sub>3</sub>		85 $\pm$ 12
UOS-28564	CH <sub>3</sub>	CH <sub>3</sub>		> 500
UOS-28565	CH <sub>3</sub>	CH <sub>3</sub>		33 $\pm$ 17
UOS-28514	CH <sub>3</sub>	CH <sub>3</sub>		> 500
UOS-28646	CH <sub>3</sub>	CH <sub>3</sub>		49 $\pm$ 5
UOS-28438	C <sub>2</sub> H <sub>5</sub>	CH <sub>3</sub>		18 $\pm$ 4
UOS-28451	CH <sub>3</sub>	OH		> 500
UOS-28527	CH <sub>3</sub>			> 500
UOS-28573	CH <sub>3</sub>			>500
UOS-28519	CH <sub>3</sub>			> 500



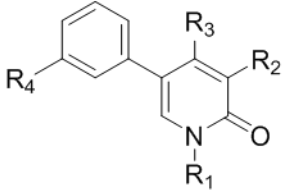
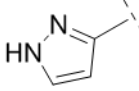
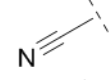

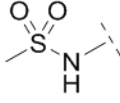
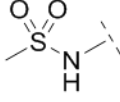
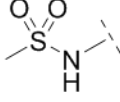
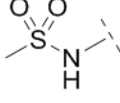
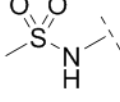
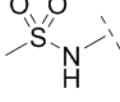
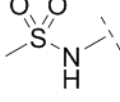
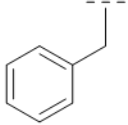
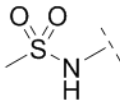
**Figure 6.30.** ITC Binding Data for Interaction Between KAT2A BRD and UOS-28438 or UOS-26378. Isotherms of raw titration heat above. Normalised binding heats below, with solid line representing nonlinear least squares fit using single-site binding model. Error calculated from deviation to least squares fit. (a) 580  $\mu\text{M}$  KAT2A BRD titrated into 60  $\mu\text{M}$  UOS-28438. (b) 600  $\mu\text{M}$  KAT2A BRD titrated into 60  $\mu\text{M}$  UOS-26378.



**Figure 6.31.** ITC Binding Data for Interaction Between KAT2A BRD and UOS-28514 or UOS-28646. Isotherms of raw titration heat above. Normalised binding heats below, with solid line representing nonlinear least squares fit using single-site binding model. Error calculated from deviation to least squares fit. (a) 580  $\mu\text{M}$  KAT2A BRD titrated into 60  $\mu\text{M}$  UOS-28514. (b) 600  $\mu\text{M}$  KAT2A BRD titrated into 60  $\mu\text{M}$  UOS-28646.

Compounds UOS-28514 and UOS-28646 were also of interest. The results of experiments conducted at ~600  $\mu\text{M}$  KAT2A BRD with 60  $\mu\text{M}$  of these compounds are shown in Figure 6.31. Reportedly, UOS-28514 exhibited reasonable potency for the PCAF BRD, with an associated  $K_d$  of 21  $\mu\text{M}$ .<sup>240</sup> However, no binding was observed by ITC in-house with the KAT2A BRD.

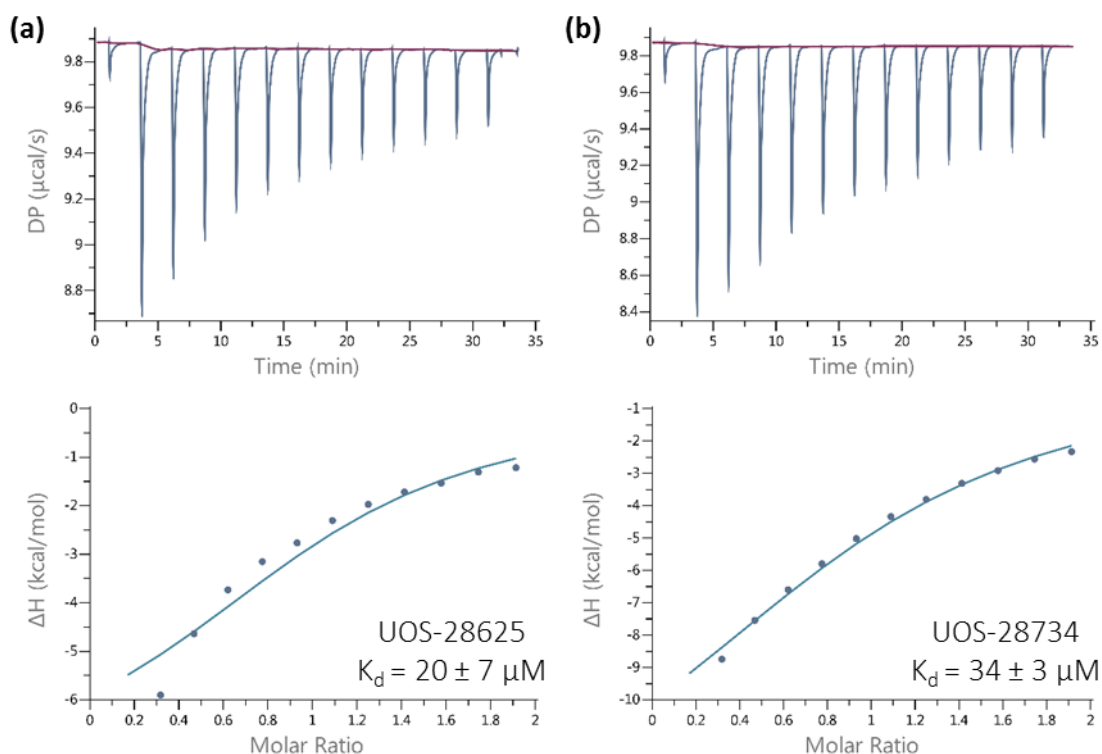
**Table 6.7.  $K_d$  Data from ITC Experiments with KAT2A BRD and Phenylpyridinone Derivatives.** Error calculated from deviation to least squares fit.

					
Compound	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	ITC $K_d$ ( $\mu\text{M}$ )
UOS-30637	CH <sub>3</sub>	CH <sub>3</sub>	H	H	34 $\pm$ 6
UOS-28642	CH <sub>3</sub>	CH <sub>3</sub>	H		36 $\pm$ 5
UOS-28643	CH <sub>3</sub>	CH <sub>3</sub>	H		144 $\pm$ 15
UOS-28734	CH <sub>3</sub>	CH <sub>3</sub>	H		34 $\pm$ 3
UOS-28625	CH <sub>3</sub>	CH <sub>3</sub>	H		20 $\pm$ 7
UOS-30731	H	CH <sub>3</sub>	H		> 500
UOS-30823	C <sub>2</sub> H <sub>5</sub>	CH <sub>3</sub>	H		76 $\pm$ 17
UOS-30655	CH <sub>3</sub>	H	H		> 500
UOS-30708	CH <sub>3</sub>	H	CH <sub>3</sub>		233 $\pm$ 102
UOS-30710	CH <sub>3</sub>	H	C <sub>2</sub> H <sub>5</sub>		350 $\pm$ 49
UOS-30754	CH <sub>3</sub>	CF <sub>3</sub>	H		476 $\pm$ 82
UOS-30714		CH <sub>3</sub>	H		148 $\pm$ 51

Inspired by the experience with SL1126 and UOS-21430, UOS-28646 was designed, with addition of a methyl group at the sulphonamide at the 7-position. UOS-28646 exhibited significantly improved binding with the KAT2A BRD, with a  $K_d$  of 49  $\mu\text{M}$ .

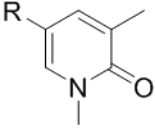
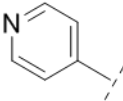
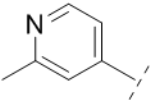
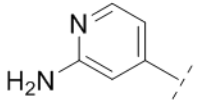
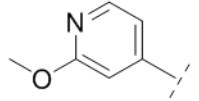
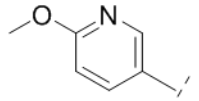
Unfortunately, despite evidence from the crystal structure of SL1126 with PCAF (PDB ID: 5FE8)<sup>240</sup> that space is available at the quinolinone 4-position, ligand growth at this vector was not tolerated. No binding was measured for compounds UOS-28527, UOS-28573 or UOS-28519.

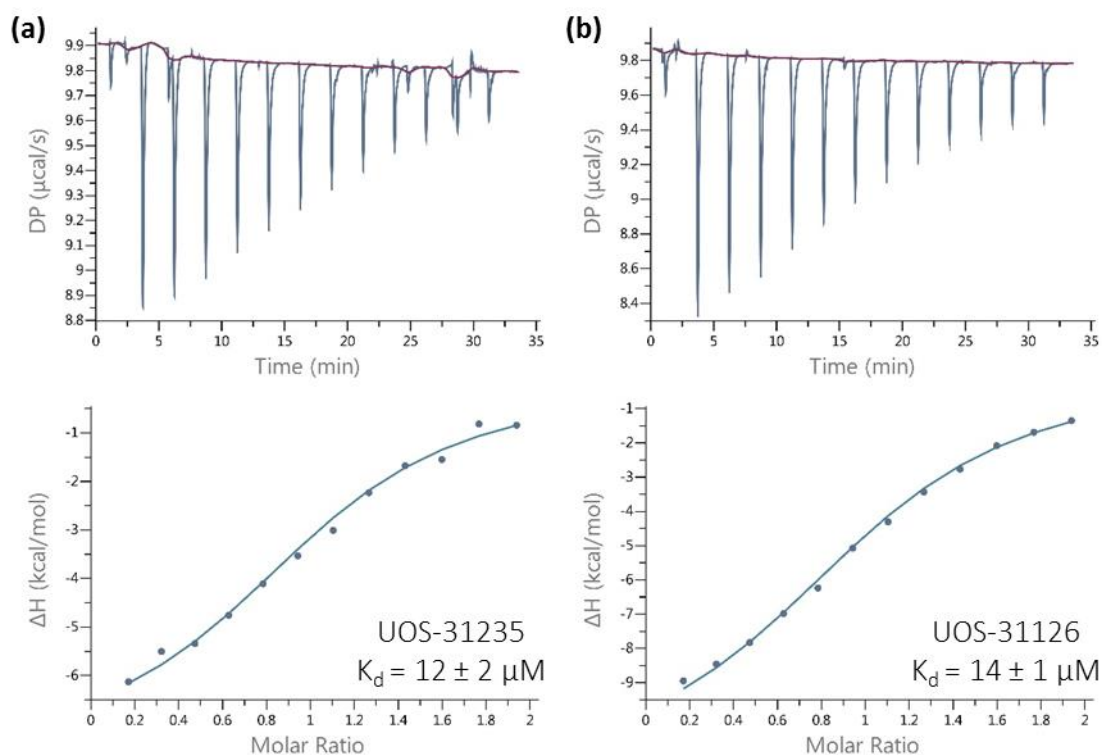
Following this, a series of phenylpyridinones, generously prepared by Dr Irina Chuckowree, Kamlesh Bala and Dr Tristan Reuillon (Sussex Drug Discovery Centre, University of Sussex), which were hoped to offer improved solubility compared to the quinolinones, were screened by ITC. Ten of the compounds exhibited binding, Table 6.7. UOS-28625 showed a small improvement in potency compared to UOS-21430 ( $K_d = 29 \mu\text{M}$ ), with a  $K_d$  of 20  $\mu\text{M}$ , but all other compounds exhibited affinity similar to or poorer than UOS-21430. The results of experiments conducted at 600  $\mu\text{M}$  KAT2A BRD with 60  $\mu\text{M}$  UOS-28625 and UOS-28734, are shown in Figure 6.32. Again, the UOS-28625 data fits poorly with the binding model, as observed with fragments UOS-18715 and UOS-28671, likely as a result of DMSO binding effects. Nonetheless, as examples of a novel scaffold, these compounds also represented interesting targets for crystallographic study.



**Figure 6.32.** ITC Binding Data for Interaction Between KAT2A BRD and UOS-28625 or UOS-28734. Isotherms of raw titration heat above. Normalised binding heats below, with solid line representing nonlinear least squares fit using single-site binding model. Error calculated from deviation to least squares fit. (a) 600  $\mu\text{M}$  KAT2A BRD titrated into 60  $\mu\text{M}$  UOS-28625. (b) 600  $\mu\text{M}$  KAT2A BRD titrated into 60  $\mu\text{M}$  UOS-28734.

**Table 6.8.  $K_d$  Data from ITC Experiments with KAT2A BRD and Pyridylpyridinone Derivatives.** Error calculated from deviation to least squares fit.

		
Compound	R	ITC $K_d$ ( $\mu$ M)
UOS-31075		$53 \pm 16$
UOS-31235		$12 \pm 2$
UOS-31234		$33 \pm 8$
UOS-31126		$14 \pm 1$
UOS-31217		$58 \pm 8$

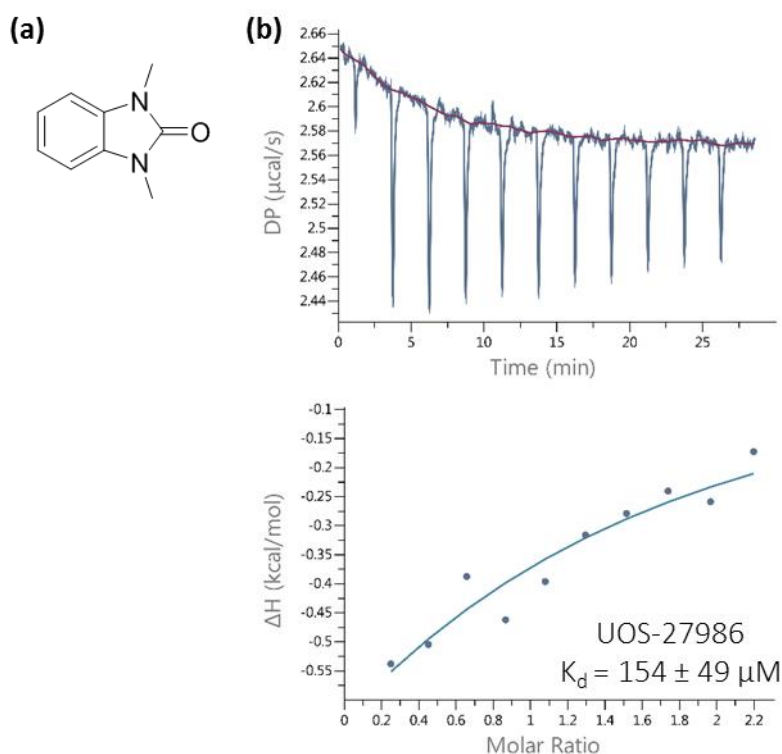


**Figure 6.33. ITC Binding Data for Interaction Between KAT2A BRD and UOS-31235 or UOS-31126.** Isotherms of raw titration heat above. Normalised binding heats below, with solid line representing nonlinear least squares fit using single-site binding model. Error calculated from deviation to least squares fit. (a) 600  $\mu$ M KAT2A BRD titrated into 60  $\mu$ M UOS-31235. (b) 600  $\mu$ M KAT2A BRD titrated into 60  $\mu$ M UOS-31126.

A final series of pyridylpyridinones, prepared by Dr Lewis Pennicott (Sussex Drug Discovery Centre, University of Sussex), were also screened by ITC. All five of these compounds exhibited

binding, Table 6.8. UOS-31235 and UOS-31126 exhibited enhanced binding compared to UOS-21430. The results of the ITC experiments with these compounds are shown in Figure 6.33.

In addition, 33 novel fragments, generously prepared by Dr Lewis Pennicott (Sussex Drug Discovery Centre, University of Sussex), were screened. KAT2A BRD binding was successfully observed for fragment UOS-27986, Figure 6.34 (a), which exhibited a  $K_d$  of 154  $\mu\text{M}$ . The result of the experiment conducted at  $\sim 600 \mu\text{M}$  KAT2A BRD with 60  $\mu\text{M}$  UOS-27986 is shown in Figure 6.34 (b). As another novel scaffold, UOS-27986 was also highlighted as an interesting target for crystallographic study.



**Figure 6.34. UOS-27986 as a Novel Binder of the KAT2A BRD.** (a) Chemical structure of UOS-27986. (b) ITC binding data for interaction between KAT2A BRD and UOS-27986. 621  $\mu\text{M}$  KAT2A BRD titrated into 60  $\mu\text{M}$  UOS-27986. Isotherm of raw titration heat above. Normalised binding heat below, with solid line representing nonlinear least squares fit using single-site binding model. Error calculated from deviation to least squares fit.

#### 6.2.8 X-Ray Crystallography

Alongside ITC screening, crystallographic studies were initiated. It was hoped that if ligand-bound structures of the KAT2A BRD with small molecule micromolar inhibitors could be obtained, these would enable improved structure-based drug design.

The crystals used to obtain the apo structure of the *Homo sapiens* GCN5 bromodomain (PDB ID: 3D7C)<sup>238</sup> reported previously, were grown in 0.1 M Tris (Tris(hydroxymethyl)aminomethane), pH 8.5, with 25% PEG (polyethylene glycol) 3350. Fortunately, these conditions were effective, and

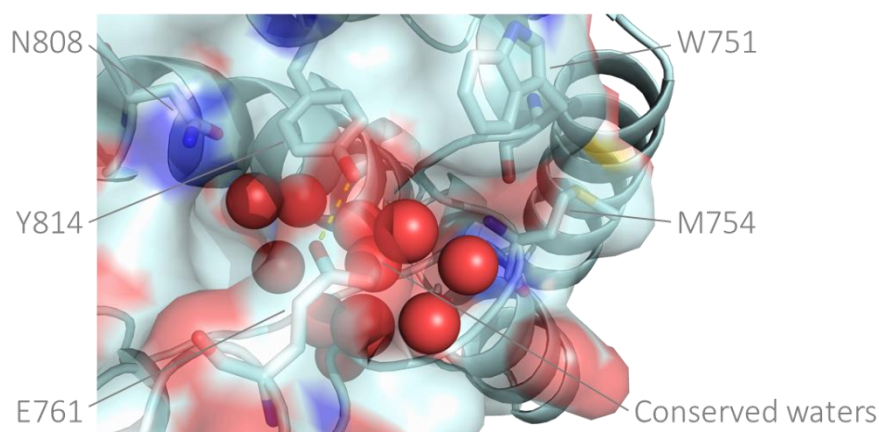


the His6-KAT2A bromodomain protein, used in the T<sub>m</sub> shift assay, was successfully crystallised in hanging drops, with 0.1 M Tris, pH 8.5 and 16% PEG 3350 at 5 mg/mL protein at 4 °C.

The resulting crystals were robust and crystal soaking, at 10 mM compound in 10% DMSO as standard, was successfully used to obtain crystals of ligand-bound KAT2A BRD, with compounds UOS-21430, UOS-18715, UOS-28438, UOS-28625, UOS-28734 and UOS-27986.

X-ray diffraction data was kindly collected using a rotating anode X-ray source (Rigaku MicroMax™-007 HF generator and Saturn 944+ CCD detector), or using synchrotron radiation at the Diamond Light Source, and the structures solved and refined by Dr Mark Roe (X-Ray Crystallography Collaborative Research Facility, University of Sussex).

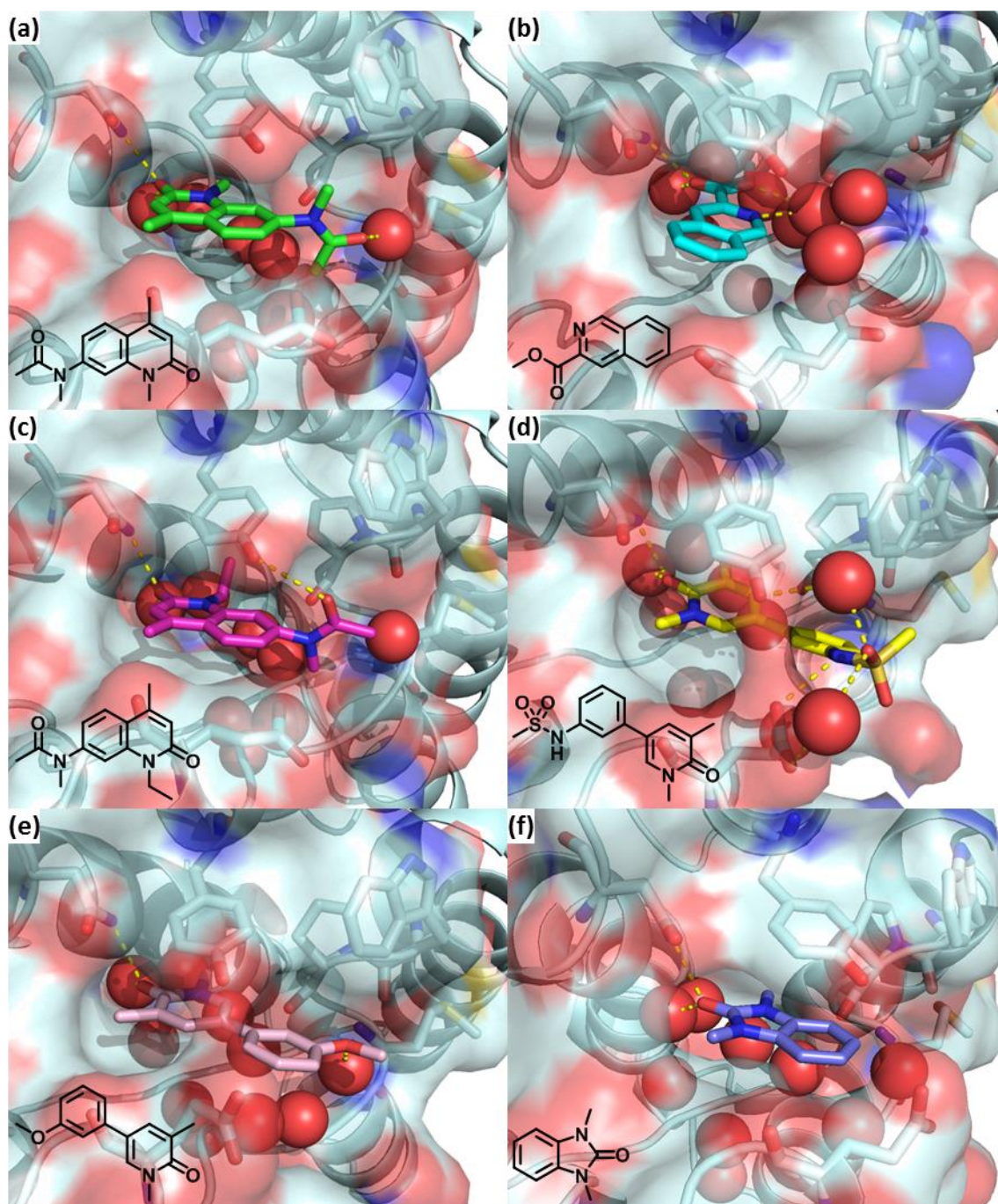
The apo structure, which was determined at a resolution of 1.53 Å, was in good agreement with the published structure of the *Homo sapiens* GCN5 bromodomain (PDB ID: 3D7C).<sup>238</sup> The druggable site, comprised of the hydrophobic acetyl-lysine binding cavity and surrounding charged surface, is depicted in Figure 6.35. The conserved asparagine N808, which interacts via hydrogen bonding with the carbonyl on acetyl-lysine residues, sits at the back of the pocket, alongside four conserved water molecules, common to BRDs. The three residues previously identified that may afford selectivity towards the KAT2A and KAT2B BRDs were of particular interest. Tryptophan W751 forms part of a lipophilic WPF shelf above the druggable pocket. Methionine M754 sits at the far right of the druggable site, distant from the hydrophobic acetyl-lysine binding cavity. Finally, glutamic acid E761 forms a hydrogen bonding interaction with tyrosine Y814, which bridges across the front of the binding site.



**Figure 6.35. Crystal Structure of Apo KAT2A BRD.** *Homo sapiens* KAT2A bromodomain. Residues either highly conserved in bromodomains or unique to KAT2A and KAT2B are displayed and annotated. Conserved waters at the back of the binding cavity and others within the pocket are displayed as red spheres.

The first ligand-bound crystal structure solved was that of the KAT2A BRD in complex with UOS-21430, shown in Figure 6.36 (a). As expected, considering the structure of the PCAF BRD with SL1126 (PDB ID: 5FE8)<sup>240</sup>, the quinolinone carbonyl at the 2-position formed a hydrogen bonding

interaction with N808 at the back of the hydrophobic acetyl-lysine binding cavity. The flat quinolinone scaffold fits neatly into the narrow BRD binding site and E761, which in the apo structure bridges across the small molecule binding site, rotates clockwise to accommodate the ligand. The increased activity of UOS-21430 compared to SL1126 appears to result from the N-methyl group causing the amide to twist relative to the quinolinone. This relieves some clashing with E761 and positions the carbonyl to interact with the solvent.



**Figure 6.36. Crystal Structures of KAT2A BRD in Complex with Small Molecules.** *Homo sapiens* KAT2A bromodomain with (a) UOS-21430, (b) UOS-18715, (c) UOS-28438, (d) UOS-28625, (e) UOS-28734 and (f) UOS-27986 bound at the acetyl-lysine binding site. Residues either highly conserved in BRDs or unique to KAT2A and KAT2B are highlighted. Conserved waters at the back of the binding cavity and others within the binding pocket are displayed as red spheres. Hydrogen bonding interactions highlighted as yellow dashed lines. 2D structures included for reference.

Attempts to generate a crystal structure of UOS-18715 in complex with the KAT2A BRD were complicated by the lower affinity of the fragment. With the standard crystal soaking protocol, using 10 mM compound in 10% DMSO, no ligand was observed at the KAT2A acetyl-lysine binding site. As DMSO is a known BRD ligand<sup>558</sup>, it was likely that at this concentration the DMSO was outcompeting the compound for KAT2A binding. Crystal soaking was therefore repeated at 100 mM compound in 20% methanol, to successfully produce a structure of the KAT2A BRD in complex with UOS-18715, Figure 6.36 (b). The bicyclic rings were rotated by 90° compared to the quinolinone UOS-21430, enabling the ester carbonyl to interact with N808 at the back of the acetyl-lysine binding pocket via a hydrogen bond, whilst avoiding clashing with E761. Again, E761 twisted away from Y814 to accommodate the ligand.

UOS-28438 was highlighted as an interesting target for crystallographic study as the ethyl group at the 1-position was expected to extend towards the conserved waters at the back of the binding cavity. As there are few reported examples in which any of these waters are displaced<sup>573</sup>, it was assumed that the binding of the ligand must be adjusted to accommodate the ethyl group. However, the crystal structure of the KAT2A BRD in complex with UOS-28438, Figure 6.36 (c), confirmed that the ethyl group can simply extend up into the hydrophobic binding pocket, pointing away from the conserved waters. The space available at this position could be tested using a series of quinolinones with propyl, isopropyl or butyl groups introduced at the 1-position.

UOS-28625 was an interesting target for crystallographic analysis as it was slightly more potent than UOS-21430 and represented a novel scaffold. The binding of UOS-28625 with KAT2A BRD, shown in Figure 6.36 (d), showed some notable differences to that of the quinolinones. As expected, the pyridinone carbonyl formed a hydrogen bonding interaction with N808 at the back of the acetyl-lysine binding site. The separation of the aromatic rings, compared to the bicyclic quinolinone scaffold, alleviated clashing with E761 and therefore the glutamic acid assumed a position more similar to that observed in the apo structure, with the carboxyl group turned to be directed away from the binding pocket. The sulphonamide moiety was then effectively placed to form hydrogen bonding interactions with E761 and with Y814 via a bridging water.

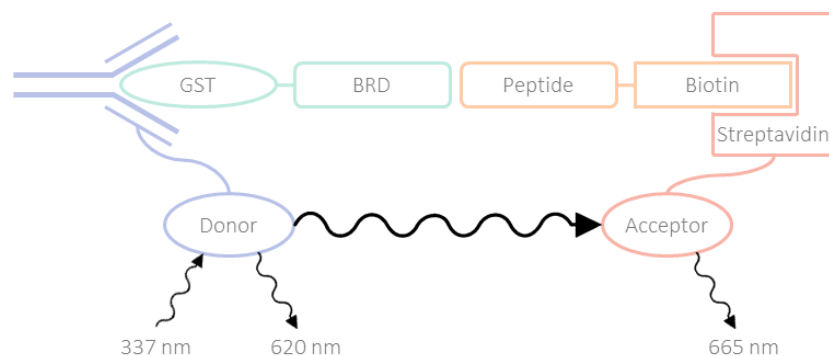
As expected, phenylpyridinone UOS-28734 interacted with the KAT2A BRD in a similar manner to UOS-28625, as shown in Figure 6.36 (e). The pyridinone carbonyl formed a hydrogen bond with N808 at the back of the hydrophobic acetyl-lysine binding cavity, and E761 assumed a position similar to that in the apo structure. The ether protruded into the pocket rather than towards the E761 carboxyl as the UOS-28625 sulphonamide did. It should be noted that for compounds UOS-28625 and UOS-28734, from the electron density maps it was not possible to

assign the orientation of the pyridinone ring regarding whether the N-methyl or C-methyl is directed towards the conserved waters at the back of the binding pocket.

Finally, the crystal structure of the KAT2A BRD in complex with UOS-27986 is shown in Figure 6.36 (f). As for UOS-18715, the low affinity of UOS-27986 meant it was more difficult to generate a ligand-bound structure. No ligand was observed at the KAT2A acetyl-lysine binding site using the standard crystal soaking protocol. Crystal soaking was repeated at 100 mM compound in 20% DMSO to successfully obtain a structure of the KAT2A BRD in complex with UOS-27986. The UOS-27986 bicycle bound in the same orientation as the quinolinone UOS-21430, the benzimidazolone carbonyl interacted with N808 at the back of the acetyl-lysine binding pocket and E761 rotated away from Y814 to enable ligand binding.

### 6.2.9 TR-FRET Assay Design

Although ITC had been utilised to generate many useful results, it was undoubtedly not well suited as a primary screening assay. ITC is time consuming, low throughput and the requirement for protein supply was unsustainably high. It was therefore essential that a new KAT2A BRD assay be developed in-house, which was better suited to high-throughput screening. Time-resolved fluorescence resonance energy transfer technology had been used to characterise inhibitors of other bromodomains such as BRD4(1) and CREBBP and p300<sup>570,571</sup>, and was therefore highlighted as a promising alternative assay.



**Figure 6.37. Schematic of Generic HTRF® Assay Design Using EPIgeneous™ Binding Domain Kit.**

Cisbio sell an assay kit, the EPIgeneous™ Binding Domain Kit, which is designed to measure the interaction between bromodomains and acetylated histone peptides, to enable characterisation of inhibitors. As shown in Figure 6.37, the interaction between Glutathione S-transferase (GST)-tagged protein and a suitable biotinylated peptide is detected using an anti-GST antibody coupled to a donor and an acceptor-labelled streptavidin (SA). When the peptide is bound to the BRD and therefore the dyes are in close proximity, the excitation of the donor triggers a FRET towards the acceptor, which then fluoresces. The HTRF® (homogeneous time resolved fluorescence) technology, refined by Cisbio, utilises cryptate donors that are excited at 337 nm

and emit at 620 nm, and red acceptors that emit at 667 nm. The signal modulates positively in proportion to the protein-peptide interaction, such that when an unlabelled small molecule inhibitor competes with the peptide for BRD binding, the fluorescence at 665 nm is reduced.

Unfortunately, KAT2A is not one of the sixteen bromodomains for which the Cisbio assay kit has been optimised. As such, there was not a recommended peptide which could be utilised in the assay. Furthermore, in peptide array screens, while the histone binding partners of several other BRDs were successfully identified, no peptides exhibited significant binding affinity for KAT2A<sup>558</sup>, and any weak binding appeared to be largely independent of peptide sequence.<sup>574</sup> However, eight peptides were identified that reportedly exhibit some affinity for the KAT2A bromodomain, including three variant HIV-1 Tat peptides acetylated at lysine 50<sup>542,575</sup>, three histone H3 peptides acetylated at lysine 14<sup>238,576</sup>, and two histone H4 peptides acetylated at lysines 12 or 16.<sup>239</sup> These peptides were purchased, along with a tetra-acetylated H4 peptide frequently employed with other BRDs in the EPIgeneous™ binding assay.

ITC was employed to measure the heats of binding of the nine peptides with the KAT2A BRD. Only peptides 35329 and 36685 showed any indication of binding with the bromodomain, Table 6.9. 35329 bound with a  $K_d$  of 67  $\mu$ M and 36685 a  $K_d$  of 282  $\mu$ M. The results of the ITC experiments conducted with KAT2A BRD and peptides 35329 and 36685 are shown in Figure 6.38. Despite representing the most potent peptide, 35329 exhibited poorer affinity than many of the small molecule inhibitors already identified and was not optimal for use in the EPIgeneous™ competition assay.

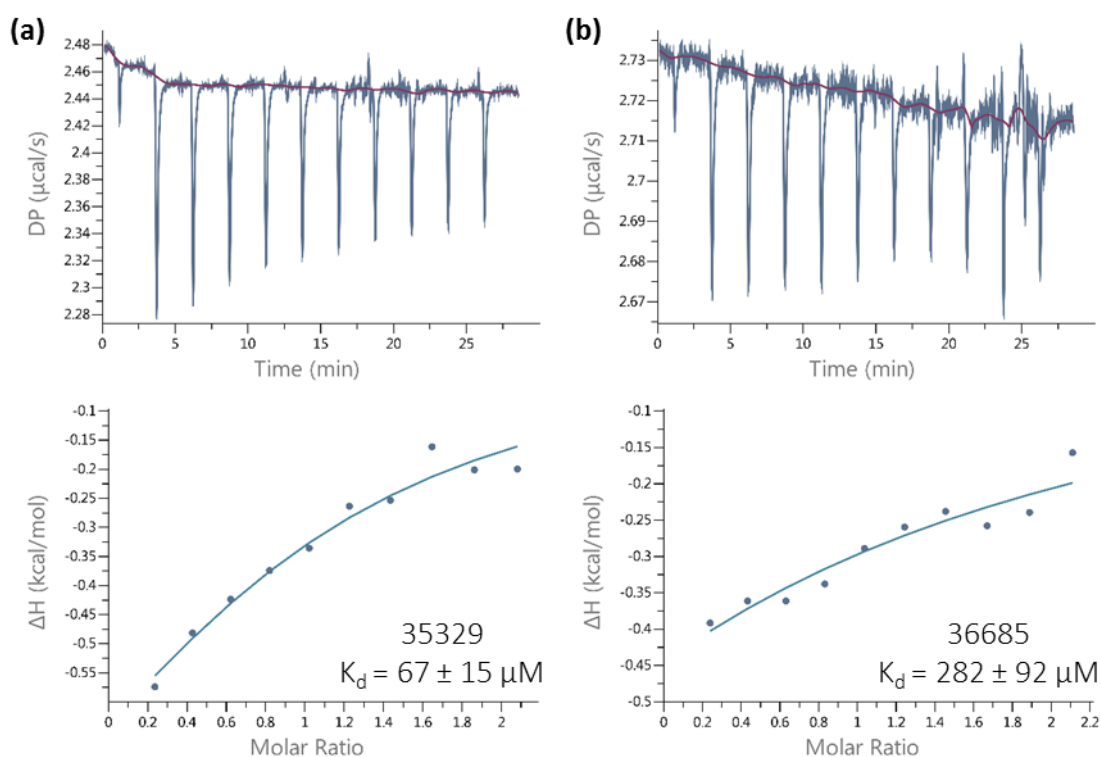
**Table 6.9.  $K_d$  Data from ITC Experiments with KAT2A BRD and Acetylated Peptides.** Error calculated from deviation to least squares fit.

Peptide	Sequence	ITC $K_d$ ( $\mu$ M)
35329	GISYGRK(Ac)KRRQRRRP	67 $\pm$ 15
36685	GISYGRK(Ac)KKRRQRRRP	282 $\pm$ 92
35330	SYGRK(Ac)KRRQR	> 500
35331	QTARKSTGGKAPRKQL	> 500
35332	QTARKSTGGK(Ac)APRKQL	> 500
35333	K(Me <sub>3</sub> )STGGK(Ac)APRK(Me <sub>3</sub> )Q	> 500
35334	LGK(Ac)GGAKRHRKVLRDNIQGI	> 500
35335	LGKGGAK(Ac)RHRKVLRDNIQGI	> 500
36684	SGRGK(Ac)GGK(Ac)GLGK(Ac)GGAK(Ac)RHRKV	> 500

At this time, a slew of PCAF and GCN5 inhibitors were reported in quick succession in patents and in the literature, Figure 6.39. First, the SGC reported several new classes of acetyl-lysine mimetic ligands with micromolar affinity for PCAF, including SL1126 as well as CPD3, which were



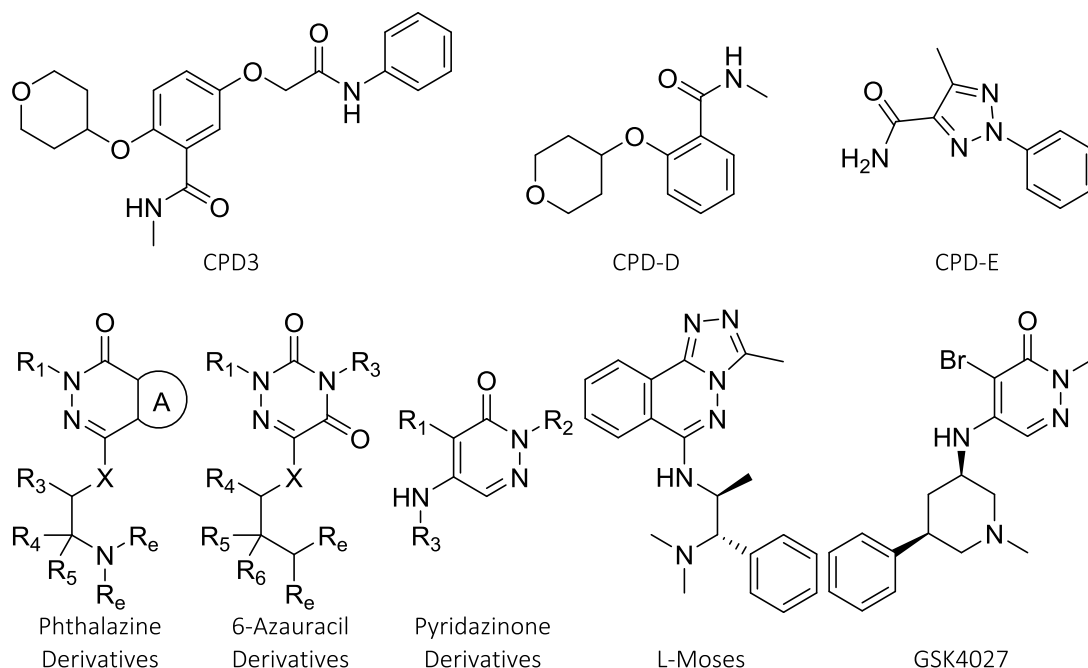
identified using thermal shift, ITC and X-ray crystallography.<sup>240</sup> Following this, a surface plasmon resonance (SPR) screen delivered fragments CPD-D and CPD-E, which exhibit 73  $\mu\text{M}$  and 250  $\mu\text{M}$  affinity for PCAF respectively.<sup>577</sup> Genentech and Constellation Pharmaceuticals then disclosed, in three patent applications, highly potent inhibitors of PCAF and GCN5, which demonstrated significant improvements in potency compared to anything reported previously. The patents covered three chemotypes, phthalazines<sup>578</sup>, 6-azauracil derivatives<sup>579</sup> and pyridazinones<sup>580</sup>. Moustakim *et al.* utilised thermal shift, ITC and X-ray crystallography data in the rational design of triazolophthalazine L-Moses, which exhibited a  $K_d$  of 126 nM for PCAF.<sup>581</sup> Finally, in January 2017, Humphreys *et al.* from GlaxoSmithKline reported GSK4027 as a dual PCAF/GCN5 chemical probe, with low nanomolar potency and  $\geq 70$ -fold selectivity over all other bromodomains.<sup>582</sup>



**Figure 6.38. ITC Binding Data for Interaction Between KAT2A BRD and Peptides 35329 and 36685.** Isotherms of raw titration heat above. Normalised binding heats below, with solid line representing nonlinear least squares fit using single-site binding model. Error calculated from deviation to least squares fit. (a) 588  $\mu\text{M}$  KAT2A BRD titrated into 60  $\mu\text{M}$  35329. (b) 497  $\mu\text{M}$  KAT2A BRD titrated into 50  $\mu\text{M}$  36685.

That the first tool inhibitors of KAT2A and KAT2B were published elsewhere was disappointing. However, the publications were highly useful. Firstly, in optimising SPR as a fragment screening tool for BRDs, Navratilova *et al.* demonstrated that DMSO interferes with BRD-inhibitor binding<sup>577</sup>, as previously noted in this work. The binding affinities of CPD-D and CPD-E for PCAF were measured at DMSO concentrations between 0% and 3%. The best results were obtained at 0% DMSO, with the apparent affinity of the compounds decreasing approximately 2.5 fold at 1% DMSO and 5 fold at 3% DMSO. As the majority of fragment screening campaigns use DMSO

as the standard solvent for preparing fragment libraries, as a compromise it was recommended that 1% DMSO should be utilised in fragment screening.<sup>577</sup> Additionally, these novel inhibitors afforded the opportunity to develop an effective KAT2A TR-FRET assay in-house.



**Figure 6.39. Recently Published KAT2A and KAT2B Binders.**

Seven inhibitors and associated enantiomers were kindly prepared by Dr Tristan Reuillon and Dr Lewis Pennicott (Sussex Drug Discovery Centre, University of Sussex), or provided by the SGC, for screening by ITC. All of the compounds except the inactive enantiomer of L-Moses (D-Moses) exhibited binding with the KAT2A bromodomain, Table 6.10. The results of the ITC experiments with the two most potent inhibitors, UOS-31388 and UOS-31181 are shown in Figure 6.40. UOS-31388 exhibited a  $K_d$  of 0.94  $\mu$ M and UOS-31181 a  $K_d$  of 1.5  $\mu$ M.

In attempt to obtain X-ray crystal structures of compounds UOS-31388 and UOS-31181 in complex with the KAT2A BRD, His6-KAT2A bromodomain protein was crystallised using 0.1 M Tris, pH 8.5 and 16% PEG 3350, in hanging drops, at 5 mg/mL protein at 4 °C. Crystals of UOS-31388 complexed with the KAT2A BRD were successfully generated using crystal soaking at 10 mM compound in 10% DMSO.

Unfortunately, crystals of UOS-31181-bound KAT2A BRD could not be obtained using the standard crystal soaking protocol. It was hypothesised that the larger UOS-31181 could not successfully access the binding site by diffusing through the solvent channels. Instead, UOS-31181 was co-crystallised with the His6-KAT2A BRD. Four 96-condition crystallisation screens were conducted with the JCSG-*plus*<sup>™</sup> HT-96, ProPlex HT-96, Morpheus<sup>™</sup> and PACT *premier*<sup>™</sup> HT-96 commercial libraries, to identify suitable crystallisation conditions. Ligand-bound His6-KAT2A

BRD was successfully crystallised in sitting drops, with 0.1 M sodium acetate, pH 5.0 and 1.5 M ammonium sulphate at 10 mg/mL protein at 20 °C.

**Table 6.10.  $K_d$  Data from ITC Experiments with KAT2A BRD and Published Binders.** Error calculated from deviation to least squares fit.

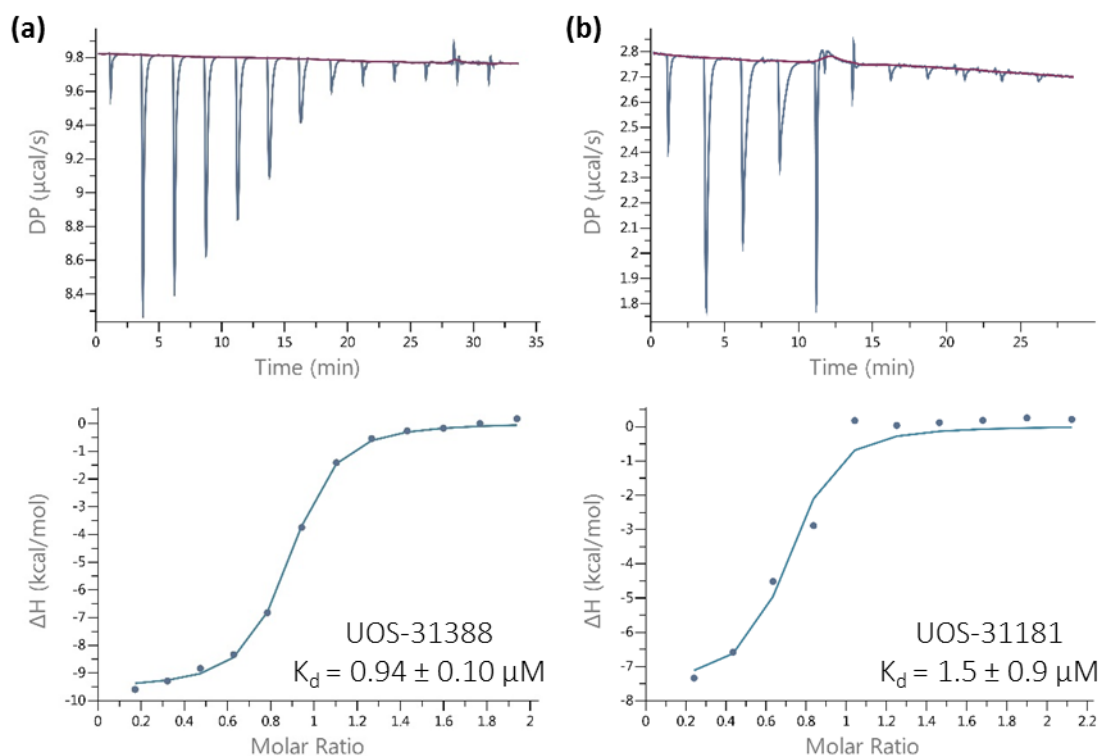
Compound	ITC $K_d$ ( $\mu$ M)
CPD3	$292 \pm 35$
L-Moses	$2 \pm 1$
D-Moses	$> 500$
UOS-31181	$1.5 \pm 0.9$
UOS-31182	$24 \pm 9$
UOS-31388	$0.94 \pm 0.10$
UOS-31389	$3.9 \pm 0.9$

X-ray diffraction data was collected using a rotating anode X-ray source (Rigaku MicroMax™-007 HF generator and Saturn 944+ CCD detector) and the structures solved and refined by Dr Mark Roe (X-Ray Crystallography Collaborative Research Facility, University of Sussex).

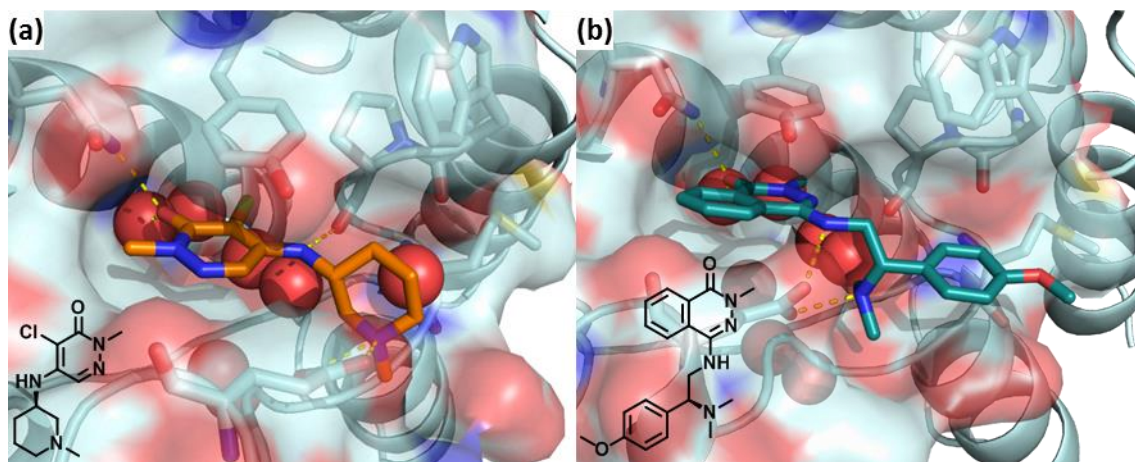
The crystal structure of UOS-31388 in complex with the KAT2A BRD was determined at a resolution of 2.20 Å. UOS-31388 exhibits remarkable shape and structural complementarity with



the BRD binding site, as shown in Figure 6.41 (a). The pyridazinone carbonyl forms a hydrogen bonding interaction with N808 at the back of the hydrophobic acetyl-lysine binding site. The amine spacer between the two rings not only alleviates clashing with E761 but also engages in a hydrogen bond with the protein backbone at proline P752. Finally, the glutamic acid E761 is drawn in, towards the small molecule, to form hydrogen bonding interactions with the piperidine nitrogen.



**Figure 6.40. ITC Binding Data for Interaction Between KAT2A BRD and UOS-31388 or UOS-31181.** Isotherms of raw titration heat above. Normalised binding heats below, with solid line representing nonlinear least squares fit using single-site binding model. Error calculated from deviation to least squares fit. (a) 600  $\mu\text{M}$  KAT2A BRD titrated into 60  $\mu\text{M}$  UOS-31388. (b) 500  $\mu\text{M}$  KAT2A BRD titrated into 50  $\mu\text{M}$  UOS-31181.

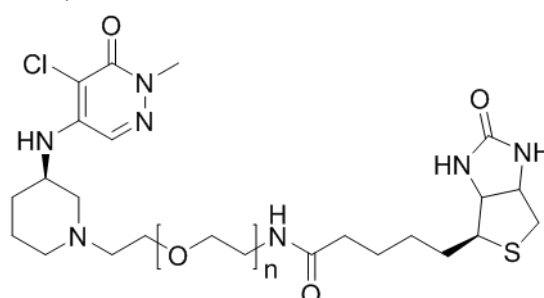


**Figure 6.41. Crystal Structures of KAT2A BRD in Complex with Published Binders.** *Homo sapiens* KAT2A BRD with (a) UOS-31388<sup>580</sup> and (b) UOS-31181<sup>578</sup> bound at the acetyl-lysine binding site. Residues either highly conserved in BRDs or unique to KAT2A and KAT2B are highlighted. Conserved waters at the back of the binding cavity and others within the binding pocket are displayed as red spheres. Hydrogen bonding interactions highlighted as yellow dashed lines. 2D structures included for reference.

The structure of UOS-31181 with KAT2A BRD was determined at a resolution of 1.80 Å. Like UOS-31388, UOS-31181 is well adapted to interact at the KAT2A BRD, Figure 6.41 (b). The phthalazinone carbonyl interacts with N808 at the back of the acetyl-lysine binding pocket via hydrogen bonding. The bicycle is rotated 90° compared to the quinolinone scaffold analysed previously, as observed for UOS-18715. The spacer between the two rings again alleviates clashing with E761, while the two amines engage in hydrogen bonding interactions with this glutamic acid. Finally, the aromatic ring sits beneath the lipophilic WPF shelf.

As the more synthetically straightforward of the two compounds, UOS-31388 was selected for use in developing an HTRF® assay in-house. It was anticipated that this small molecule inhibitor could substitute the peptide required in the competition assay. Dr Lewis Pennicott (Sussex Drug Discovery Centre, University of Sussex) kindly synthesised compound UOS-31490, Table 6.11, an analogue of UOS-31388 joined via a PEG linker to biotin, at the solvent accessible N-methyl position. ITC was employed to measure the heat of binding of UOS-31490 with KAT2A. The result of the experiment at 600 µM protein and 60 µM compound is shown in Figure 6.42 (a). Despite the significant increase in size, UOS-31490 retained good affinity, exhibiting a  $K_d$  of 4.9 µM.

**Table 6.11.  $K_d$  Data from ITC Experiments with KAT2A BRD and Prospective Ligand-Biotin Peptide Substitutes.** Error calculated from deviation to least squares fit.

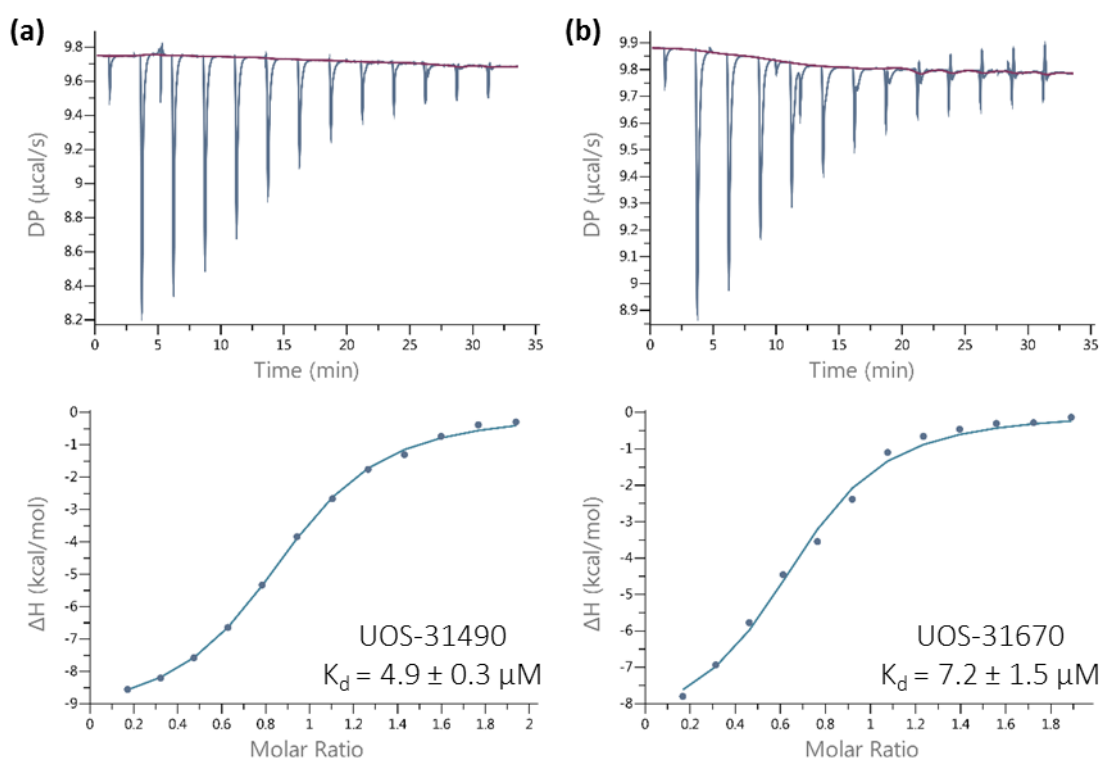


Compound	n	ITC $K_d$ (µM)
UOS-31490	1	4.9 ± 0.3
UOS-31669	2	7.7 ± 1.4
UOS-31670	3	7.2 ± 1.5

Having finally attained a biotinylated ligand with low micromolar affinity for the KAT2A BRD, efforts began to establish a TR-FRET assay in-house. N-terminal GST-tagged human KAT2A bromodomain and the EPIgeneous™ Binding Domain Discovery Kit were purchased from Cisbio. This kit is designed to determine the combination of donor and acceptor dyes best suited to measuring the interaction between the target protein and biotinylated ligand.

Firstly, ligand-biotin titrations were conducted with 0.1 nM to 30000 nM UOS-31490, at fixed concentrations of 5 nM GST-KAT2A and anti-GST-donor cryptate, with a constant ligand-biotin : streptavidin-acceptor ratio of 8:1, as recommended. Negative controls, lacking GST-KAT2A, were

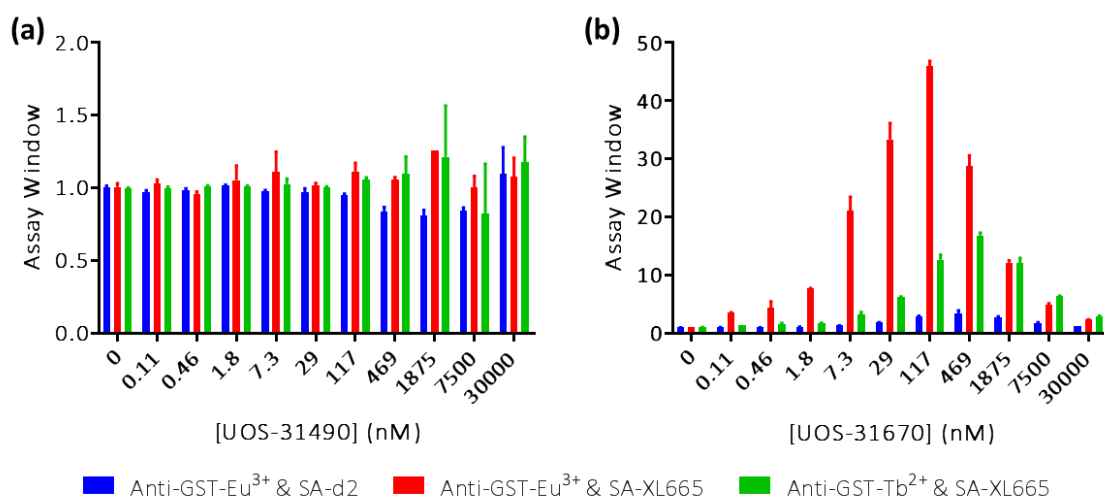
included to assess the non-specific signal at each concentration of UOS-31490. These were used to calculate the assay window (= HTRF<sup>®</sup> ratio positive / HTRF<sup>®</sup> ratio negative). The fluorescence emission was read at wavelengths 665 nm and 620 nm and the results expressed in HTRF<sup>®</sup> ratio (= (665 nm/620 nm)  $\times 10^4$ ). Three detector reagent pairings were assessed, anti-GST-Eu<sup>3+</sup> with SA-d2, anti-GST-Eu<sup>3+</sup> with SA-XL665 and anti-GST-Tb<sup>2+</sup> with SA-XL665. The resulting assay windows are shown in Figure 6.43 (a).



**Figure 6.42. ITC Binding Data for Interaction Between KAT2A BRD and UOS-31490 or UOS-31670.** Isotherms of raw titration heat above. Normalised binding heats below, with solid line representing nonlinear least squares fit using single-site binding model. Error calculated from deviation to least squares fit. (a) 600  $\mu\text{M}$  KAT2A BRD titrated into 60  $\mu\text{M}$  UOS-31490. (b) 585  $\mu\text{M}$  KAT2A BRD titrated into 60  $\mu\text{M}$  UOS-31670.

Unfortunately, no assay window was observed with any detector reagent pair at any concentration of UOS-31490. It was hypothesised that the linker between the UOS-31388 and biotin moieties may not enable sufficient flexibility, such that the biotin is shrouded by the protein and unable to interact with the streptavidin-acceptor. It was anticipated that use of a longer linker chain, comprised of multiple PEGs, would liberate the biotin moiety.

Dr Lewis Pennicott (Sussex Drug Discovery Centre, University of Sussex) synthesised two UOS-31490 analogues with increased linker length. ITC was employed to measure the heats of binding of these compounds, UOS-31669 and UOS-31670, with the KAT2A BRD. Both UOS-31669 and UOS-31670 retained good affinity, with  $K_d$ s of 7.7  $\mu\text{M}$  and 7.2  $\mu\text{M}$  respectively, Table 6.11. UOS-31670, which offered the longest linker, was selected for use in the TR-FRET assay. The result of the experiment conducted at 600  $\mu\text{M}$  protein and 60  $\mu\text{M}$  UOS-31670 is shown in Figure 6.42 (b).



**Figure 6.43. KAT2A BRD TR-FRET Assay Design Using EPIgeneous™ Binding Domain Discovery Kit.** Titrations of 0.11 nM to 30000 nM ligand-biotin and SA-acceptor (SA-d2 or SA-XL665) at a ligand-biotin : streptavidin-acceptor ratio of 8:1, added to 5 nM GST-KAT2A BRD with 1x anti-GST-donor cryptate ( $\text{Eu}^{3+}$  or  $\text{Tb}^{2+}$ ), 20  $\mu\text{L}$  well volume, and incubated for three hours. HTRF® ratios calculated and compared with negative controls without GST-KAT2A to determine assay windows. Optimised with ligand-biotin species (a) UOS-31490 and (b) UOS-31670. Graphs represent mean  $\pm$  1 SD of duplicate repeats.

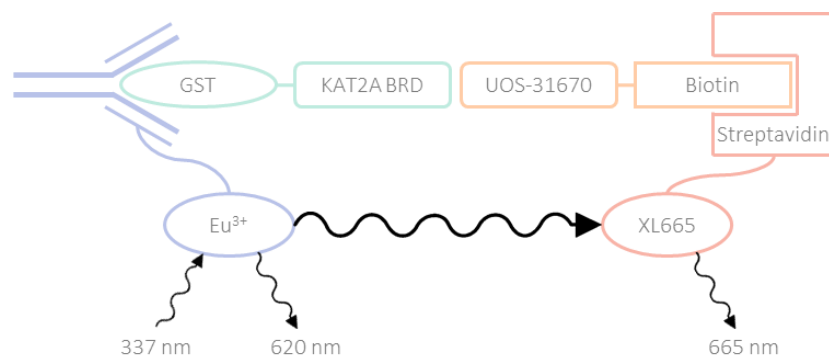
The ligand-biotin titrations were repeated with UOS-31670, again utilising 0.1 nM to 30000 nM ligand-biotin, fixed concentrations of 5 nM GST-KAT2A and anti-GST-donor cryptate, a constant UOS-31670 : streptavidin-acceptor ratio of 8:1, and including negative controls without GST-KAT2A. The fluorescence emission was read at wavelengths 665 nm and 620 nm to attain the HTRF® ratio and the assay windows were calculated. The same three detector reagent pairings were assessed. The derived assay windows are shown in Figure 6.43 (b).

Gratifyingly, greatly increased assay windows were observed. The best HTRF® conjugate combination was anti-GST- $\text{Eu}^{3+}$  with SA-XL665, which gave the largest assay windows and would therefore afford the most robust assay. This detector reagent pair was therefore selected for assay optimisation and the corresponding reagents and buffers were purchased as appropriate.

#### 6.2.10 TR-FRET Optimisation and Screening

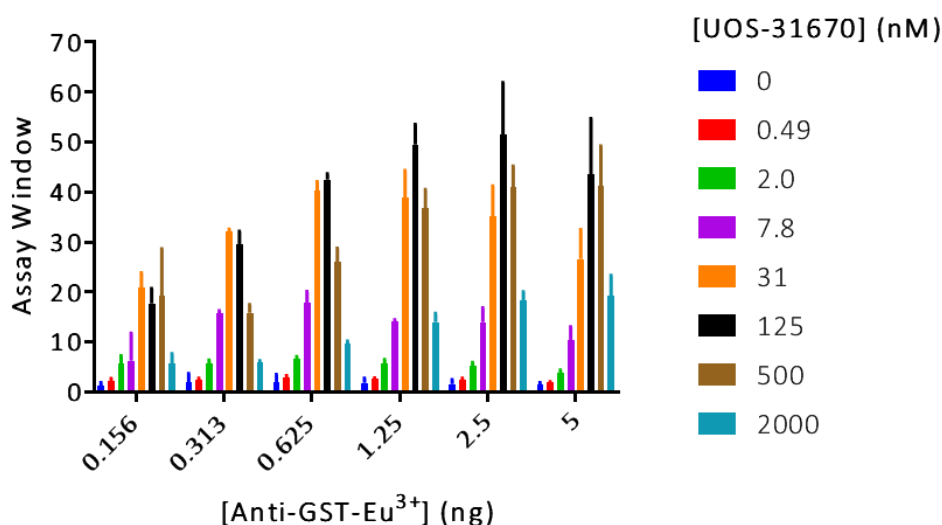
These results fundamentally defined the design for the in-house TR-FRET assay. As discussed, Cisbio EPIgeneous™ Binding Domain assays are designed to measure the interaction between bromodomains and suitable peptides, to enable characterisation of inhibitors. In this assay however, UOS-31670 was used in lieu of a suitable peptide. The interaction between GST-KAT2A BRD and UOS-31670 was detected using an anti-GST antibody coupled to an  $\text{Eu}^{3+}$  cryptate donor and an XL665 acceptor-labelled streptavidin, Figure 6.44. The  $\text{Eu}^{3+}$  cryptate is excited at 337 nm and emits at 620 nm. If UOS-31670 is bound to the GST-KAT2A BRD, the dyes are close and the excitation of the cryptate triggers a FRET towards XL665, which fluoresces at 665 nm. The specific signal, HTRF®  $\Delta$  ratio (= HTRF® positive - HTRF® negative), modulates in proportion with

the KAT2A BRD-UOS-31670 interaction. Binding of unlabelled small molecules is measured via competition with the ligand-biotin, UOS-31670.



**Figure 6.44. Schematic of In-House HTRF® Assay Design Using EPIgeneous™ Binding Domain Kit.**

It was anticipated that post-optimisation this TR-FRET assay would be utilised to conduct a fragment screen in-house. Accordingly, it was decided that 1% DMSO should be utilised as standard in the assay, as recommended by Navratilova *et al.* in 2016.<sup>577</sup> As a known BRD inhibitor, DMSO was expected to influence the HTRF® signal.<sup>558</sup> As such, all optimisation experiments were conducted in the presence of 1% DMSO.

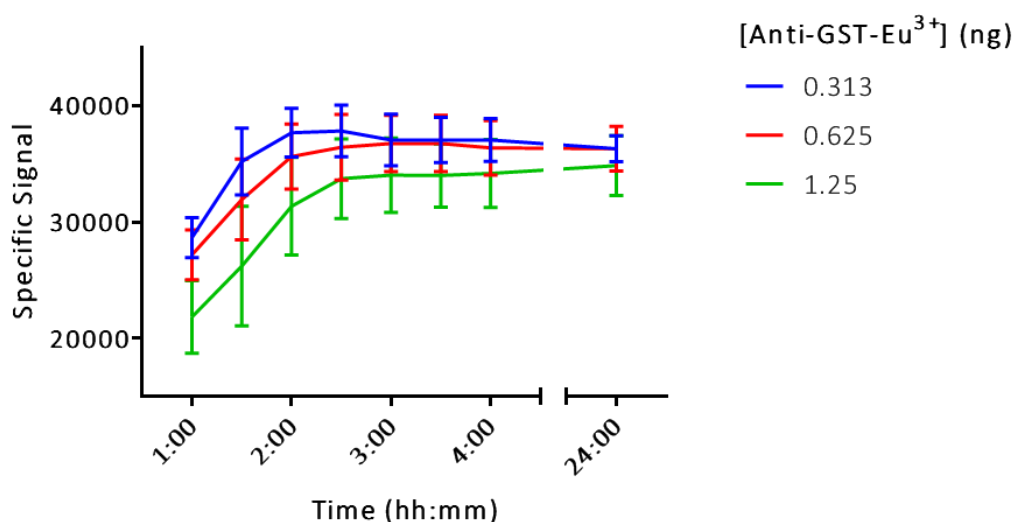


**Figure 6.45. KAT2A BRD TR-FRET Assay Anti-GST-Eu³⁺ Optimisation.** Titrations of 0.156 ng to 5 ng anti-GST-Eu³⁺ and 0.49 nM to 2000 nM UOS-31670 with SA-XL665 at a ligand-biotin : streptavidin-acceptor ratio of 8:1, added to 5 nM GST-KAT2A BRD, 20 µL well volume, and incubated for two hours. HTRF® ratios calculated and compared with negative controls without GST-KAT2A to determine assay windows. Graph represents mean ± 1 SD of two experiments.

In the EPIgeneous™ Binding Domain Discovery Kit the anti-GST-Eu³⁺ reagent was supplied at 50x working concentration. However, when purchased at higher quantities for assay optimisation and screening it was provided as a powder. As a result, assay optimisation necessarily began with establishing the approximate required concentration of the Eu³⁺ cryptate. 5 nM GST-KAT2A was incubated for two hours with 0.156 ng to 5 ng anti-GST-donor cryptate and 0.49 nM to

2000 nM ligand-biotin with a constant UOS-31670 : streptavidin-acceptor ratio of 8:1. Negative controls without GST-KAT2A were included. The fluorescence emission was read at wavelengths 665 nm and 620 nm to attain the HTRF® ratio and the assay windows were calculated, Figure 6.45. At every concentration of anti-GST-Eu<sup>3+</sup>, assay windows of greater than 20x were observed with appropriate concentrations of UOS-31670. 0.313 ng to 1.25 ng anti-GST-Eu<sup>3+</sup> afforded good assay windows with small associated error. Under these conditions, UOS-31670 concentrations of between 7.8 nM and 500 nM were optimal.

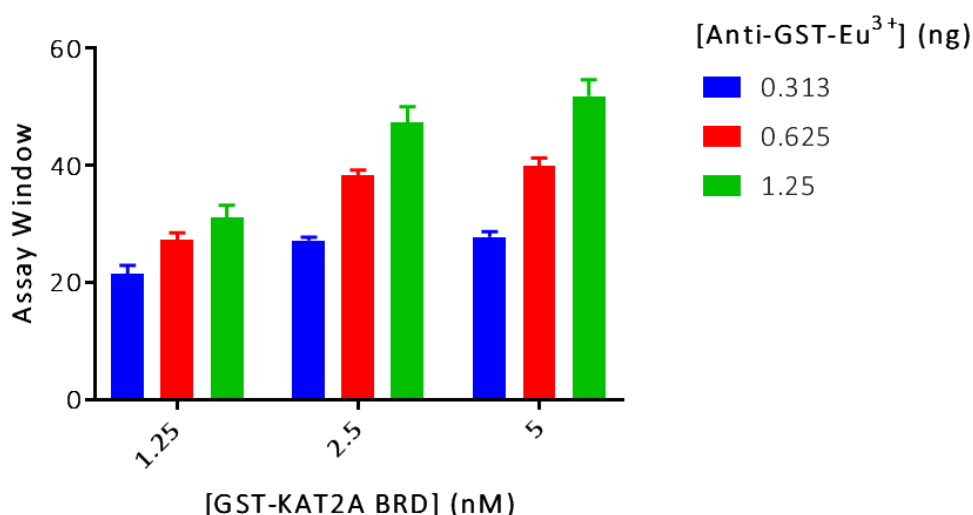
During the aforementioned primary experiments it was noticed that the fluorescence signal attributed to the GST-KAT2A BRD interacting with UOS-31670 exhibits some variation over time. To determine the impact of this fluctuation on the specific signal it was necessary to monitor the fluorescence over a series of incubation periods. 5 nM GST-KAT2A was incubated for between 1 and 24 hours with 0.313 ng to 1.25 ng anti-GST-donor cryptate and 125 nM UOS-31670 with 8:1 ligand-biotin : streptavidin-acceptor, with negative controls. The fluorescence emission was read at 665 nm and 620 nm to attain the HTRF® ratios and the specific signals were calculated, Figure 6.46. The specific signal developed gradually during early incubation but after two hours exhibited no further significant increase with time. It was therefore concluded that assay plates should be incubated for at least two hours to ensure that the specific signal is maximised and to achieve consistency between plates.



**Figure 6.46. KAT2A BRD TR-FRET Assay Incubation Period Optimisation.** Titration of 0.313 ng to 1.25 ng anti-GST-Eu<sup>3+</sup> added to 125 nM UOS-31670 with 15.6 nM SA-XL665 and 5 nM GST-KAT2A BRD, 20  $\mu$ L well volume, and incubated for 1-24 hours. HTRF® ratios calculated and the background signal, deduced from negative controls lacking GST-KAT2A, subtracted to determine the specific signal. Graph represents mean  $\pm$  1 SD of two experiments.

As the N-terminal GST-tagged human KAT2A bromodomain protein was purchased from Cisbio, there was a financial incentive to employ a low protein concentration in the assay. 1.25 nM to 5 nM GST-KAT2A was incubated for two hours with 0.313 ng to 1.25 ng anti-GST-Eu<sup>3+</sup> and

200 nM UOS-31670 with 8:1 ligand-biotin : streptavidin-acceptor, including negative controls. The fluorescence emission was read at wavelengths 665 nm and 620 nm to attain the HTRF® ratio and the assay windows were calculated, Figure 6.47. There was no significant reduction in assay window using 2.5 nM GST-KAT2A BRD compared to 5 nM.



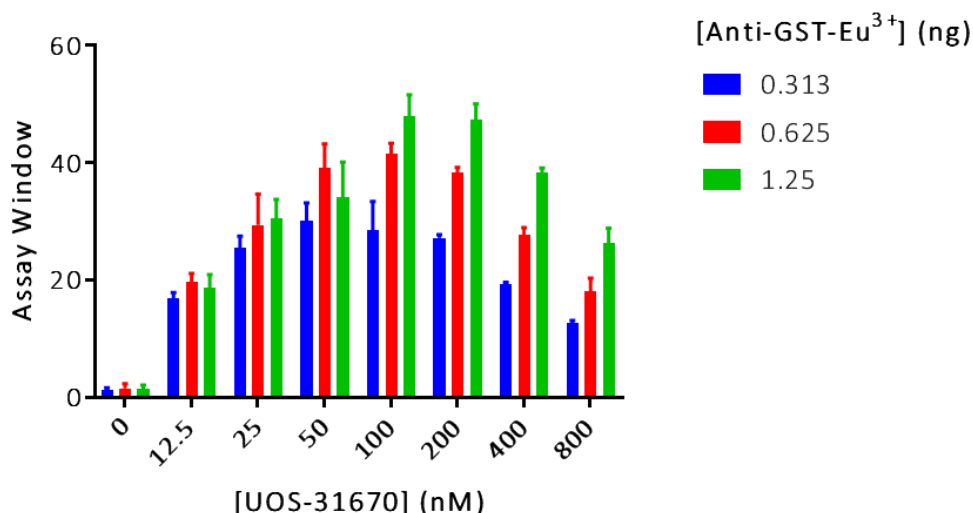
**Figure 6.47. KAT2A BRD TR-FRET Assay Protein Concentration Optimisation.** Titration of 0.313 ng to 1.25 ng anti-GST-Eu<sup>3+</sup> added to 200 nM UOS-31670 with 25 nM SA-XL665 and 1.25 nM to 5 nM GST-KAT2A BRD, 20  $\mu$ L well volume, and incubated for two hours. HTRF® ratios calculated and compared with negative controls without GST-KAT2A to determine assay windows. Graph represents mean  $\pm$  1 SD of two experiments.

Following this, it was necessary to determine the most appropriate combination of concentrations of the HTRF® conjugates. A fixed concentration of 2.5 nM GST-KAT2A was incubated with 0.313 ng to 1.25 ng anti-GST-donor cryptate and 12.5 nM to 800 nM UOS-31670 with 8:1 ligand-biotin : streptavidin-acceptor, including negative controls as standard. The fluorescence emission was measured at 665 nm and 620 nm to calculate the HTRF® ratio and the assay windows were calculated, Figure 6.48. Again, all concentrations of anti-GST-Eu<sup>3+</sup> afforded assay windows of greater than 20x when combined with appropriate UOS-31670 concentrations. With 0.313 ng anti-GST-Eu<sup>3+</sup> the largest assay window was observed at between 50 nM and 100 nM UOS-31670, with 0.625 ng anti-GST-Eu<sup>3+</sup> at roughly 100 nM and with 1.25 ng at between 100 nM and 200 nM UOS-31670.

Similarly to the GST-KAT2A BRD protein, anti-GST-Eu<sup>3+</sup> and SA-XL665 are commercial reagents, and therefore there is an economic incentive to use lower concentrations. Plate tests were conducted, in which 384 well assay plates were filled with 192 low and 192 high controls, to test the reliability of automation and expose any edge effects. 2.5 nM GST-KAT2A was incubated with 0.313 ng anti-GST-Eu<sup>3+</sup> and 80 nM UOS-31670, or 0.625 ng or 1.25 ng anti-GST-Eu<sup>3+</sup> and 120 nM UOS-31670, with 8:1 UOS-31670 : streptavidin-acceptor. The fluorescence emission was read at wavelengths 665 nm and 620 nm to attain the HTRF® ratios and the Z-factor was



calculated for each of the three plates as a measure of statistical effect size, Table 6.12. All three HTRF® conjugate concentration pairings exhibited good Z'. It was therefore concluded that although higher anti-GST-donor cryptate concentrations might afford a slightly increased assay window, Figure 6.48, to reduce the cost of the assay and therefore optimise for HTS, 0.313 ng anti-GST-Eu<sup>3+</sup>, 80 nM UOS-31670 and 10 nM SA-XL665 would be employed as standard.



**Figure 6.48. KAT2A BRD TR-FRET Assay UOS-31670 and Anti-GST-Eu<sup>3+</sup> Optimisation.** Titrations of 0.313 ng to 1.25 ng anti-GST-Eu<sup>3+</sup> and 12.5 nM to 800 nM UOS-31670 with SA-XL665 at a ligand-biotin : streptavidin-acceptor ratio of 8:1, added to 2.5 nM GST-KAT2A BRD, 20  $\mu$ L well volume, and incubated for two hours. HTRF® ratios calculated and compared with negative controls lacking GST-KAT2A to determine assay windows. Graph represents mean  $\pm$  1 SD of two experiments.

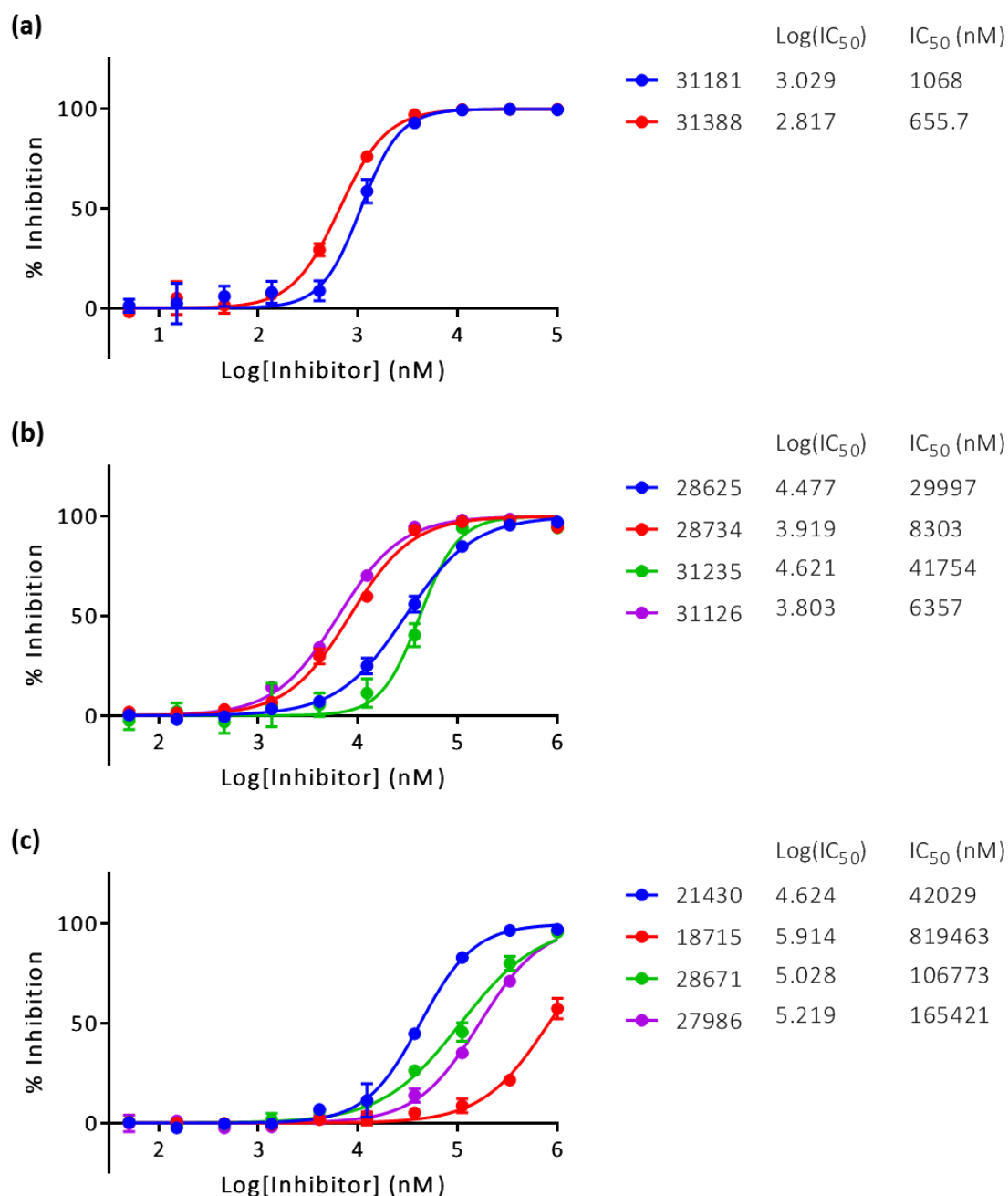
**Table 6.12. Assessing Reliability of TR-FRET Assay with Automation.** 384 well assay plates filled with 50% high and 50% low controls. Anti-GST-Eu<sup>3+</sup>, UOS-31670 and SA-XL665 incubated for two hours with or without 2.5 nM GST-KAT2A. HTRF® ratios calculated and Z' determined to measure statistical effect size.

[Anti-GST-Eu <sup>3+</sup> ] (ng)	[UOS-31670] (nM)	[SA-XL665] (nM)	Z-Prime
1.25	120	15	0.789
0.625	120	15	0.838
0.313	80	10	0.865

To test these assay conditions, IC<sub>50</sub>s were determined for a series of ten known KAT2A BRD inhibitors, previously assessed in-house using ITC. The IC<sub>50</sub> is defined as the concentration of a competitive inhibitor that displaces 50% of the labelled ligand. The compounds were serially diluted at 1/3 to achieve a ten point dilution curve and incubated with 2.5 nM GST-KAT2A, 0.313 ng anti-GST-Eu<sup>3+</sup>, 80 nM UOS-31670 and 10 nM SA-XL665 for two hours. High (positive) and low (negative) controls without inhibitor and with or without GST-KAT2A respectively, were included to determine the maximum and minimum signal. The fluorescence emission was read at 665 nm and 620 nm to attain the HTRF® ratios, and the specific signals were calculated and converted into percentage inhibition ( $= 100 - ((\text{HTRF}^\circ \Delta \text{ratio})_{\text{sample}} / (\text{HTRF}^\circ \Delta \text{ratio})_{\text{pos ctrl}}) \times 100$ ). The percentage inhibition was plotted against log[inhibitor], Figure 6.49, and the IC<sub>50</sub>s were



determined using the nonlinear regression,  $\log[\text{inhibitor}]$  vs. normalised response – variable slope equation in GraphPad Prism 7.02.<sup>243</sup>



**Figure 6.49. Dose-Response Curves Showing Percentage Inhibition of KAT2A BRD-UOS-31670 Binding.** Serial dilution of small molecule added to 0.313 ng anti-GST-Eu<sup>3+</sup>, 80 nM UOS-31670, 10 nM SA-XL665 and 2.5 nM GST-KAT2A BRD, 20  $\mu$ L well volume, and incubated for two hours. HTRF® ratios calculated and background signal, deduced from low controls lacking GST-KAT2A, subtracted to determine specific signal. Specific signal converted to percentage inhibition dependent on high controls. (a) Ten-point dilutions of UOS-31181 and UOS-31388 serially diluted 1/3 with top concentration 100  $\mu$ M. (b) Ten-point dilutions of UOS-28625, UOS-28734, UOS-31235 and UOS-31126 serially diluted 1/3 with top concentration 1 mM. (c) Ten-point dilutions of UOS-21430, UOS-18715, UOS-28671 and UOS-27986 serially diluted 1/3 with top concentration 1 mM. Graph represents mean  $\pm$  1 SD of two experiments.

The Cheng-Prusoff equation defines the theoretical relationship between the IC<sub>50</sub> measured for a competitive inhibitor with given  $K_i$ , the concentration of the labelled ligand, [L], and the  $K_d$  of

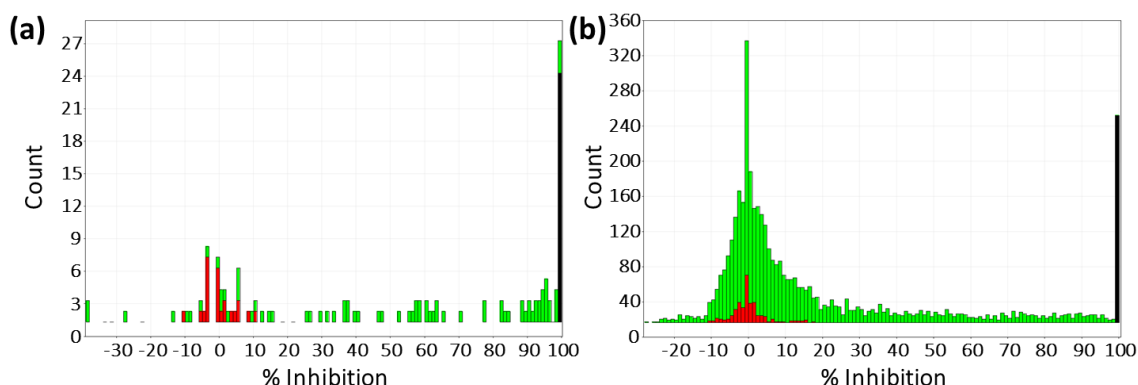
the labelled ligand-receptor interaction. The inhibition constant,  $K_i$ , of the unlabelled ligand, defined as the equilibrium concentration of the inhibitor that would occupy 50% of the receptor sites in the absence of competing labelled ligand, is calculated as  $K_i = IC_{50} / (1 + ([L]/K_d))$ . In this HTRF® assay,  $[L]$  is small compared to  $K_d$ . The UOS-31670-KAT2A BRD interaction has an associated  $K_d$  of 7.2  $\mu$ M and UOS-31670 is utilised at a concentration of just 80 nM. As a result, under these assay conditions  $K_i \approx IC_{50}$  and therefore the calculated  $IC_{50}$  values can be compared directly with the dissociation constants determined by ITC, Table 6.13. Encouragingly the  $IC_{50}$ s determined by TR-FRET compared very well with those observed by ITC. For the majority of the compounds the calculated values differed by less than two-fold and all were of the same order of magnitude, with the exception of UOS-18715. As discussed previously, there was a lack of confidence in the  $K_d$  calculated by ITC for the interaction of UOS-18715 and the KAT2A BRD, due to the low affinity of the fragment. The measured heats of binding were small, which permitted DMSO binding effects to impact the data, increasing the error in the value, and it was expected that the titration represented only a small section of the binding curve, obligating extensive data extrapolation. This could explain the considerable variation between the HTRF®  $IC_{50}$  and ITC  $K_d$ .

**Table 6.13. Comparison of  $K_d$  and  $IC_{50}$  Values determined by ITC and HTRF® Assays.**

Compound	ITC $K_d$ ( $\mu$ M)	HTRF® $IC_{50}$ ( $\mu$ M)
UOS-31181	1.5	1.07
UOS-31388	0.94	0.656
UOS-28625	20	30.0
UOS-28734	34	8.30
UOS-31235	12	41.8
UOS-31126	14	6.36
UOS-21430	29	42.0
UOS-18715	42	819
UOS-28671	117	107
UOS-27986	154	165

Having demonstrated that the  $IC_{50}$ s calculated using the TR-FRET assay compare well with the  $K_d$ s determined by ITC, a small single point screen was conducted in duplicate, to test 80 compounds available in-house, generously prepared by Dr Lewis Pennicott (Sussex Drug Discovery Centre, University of Sussex). 1 mM of each compound was incubated with 2.5 nM GST-KAT2A, 0.313 ng anti-GST-Eu<sup>3+</sup>, 80 nM UOS-31670 and 10 nM SA-XL665 for two hours. High and low controls were included to determine the maximum and minimum signal. The fluorescence emission was read at 665 nm and 620 nm and the HTRF® ratios were calculated. The specific signals were determined and converted into percentage inhibition using the Studies package from Dotmatics.<sup>244</sup>

The data was analysed using the Vortex package from Dotmatics.<sup>244</sup> The results from the first repeat are illustrated in Figure 6.50 (a). Reassuringly, the controls were tightly distributed, with the positive controls presenting standard deviations of 5.20 and 4.75 for the first and second repeats respectively. The correlation between the two repeats is displayed in Figure 6.51 (a). The repeats correlated strongly, with a Pearson correlation coefficient of 0.84, computed using GraphPad Prism 7.02<sup>243</sup>, indicative of good reproducibility of data.

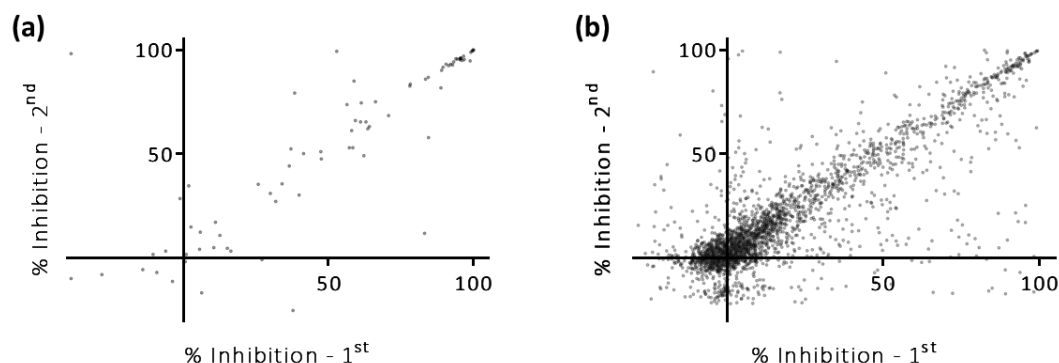


**Figure 6.50. Distribution of Percentage Inhibition Observed in Single Point Screens.** 1 mM small molecule added to 0.313 ng anti-GST-Eu<sup>3+</sup>, 80 nM UOS-31670, 10 nM SA-XL665 and 2.5 nM GST-KAT2A BRD, 20  $\mu$ L well volume, and incubated for two hours. HTRF<sup>®</sup> ratios calculated and background signal, deduced from low controls without GST-KAT2A (shown in black), subtracted to determine specific signal. Specific signal converted to percentage inhibition dependent on high controls (shown in red). (a) Screen of 80 small molecules available in-house. (b) Screen of 3140 fragments from Maybridge Ro3 2500 Diversity Fragment Library and Life Chemicals 3D Fragments Library. Graphs represent first of the two repeats.

60 compounds induced significant inhibition (over 15%) of GST-KAT2A-UOS-31670 binding, calculated as greater than three times the standard deviation in the controls, and were progressed to IC<sub>50</sub> screening, which was again conducted in duplicate. The compounds were serially diluted at 1/3 to obtain ten point dilution curves with a top concentration of 1 mM. The compound dilutions were incubated with 2.5 nM GST-KAT2A, 0.313 ng anti-GST-Eu<sup>3+</sup>, 80 nM UOS-31670 and 10 nM SA-XL665 for two hours. High and low controls were included as standard. The fluorescence emission was read at 665 nm and 620 nm and the HTRF<sup>®</sup> ratios were calculated. The specific signals were calculated and converted into percentage inhibition and the resulting IC<sub>50</sub>s were determined using the Studies package from Dotmatics.<sup>244</sup>

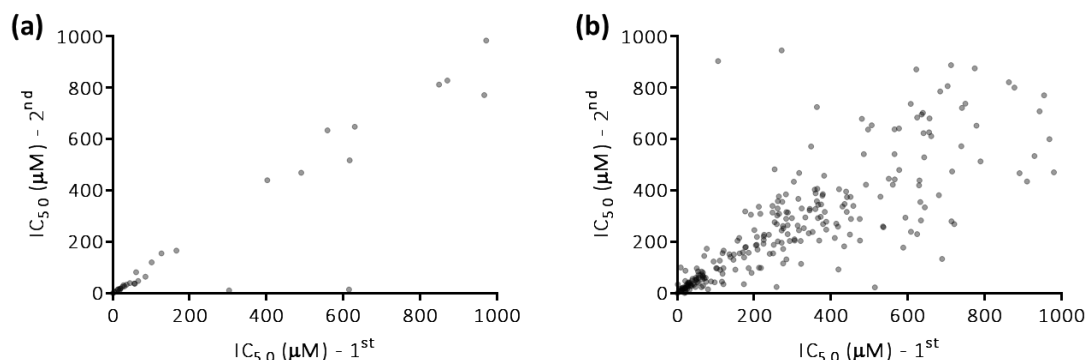
For 18 of the 60 compounds tested, the IC<sub>50</sub> could not be calculated, indicative of a value greater than 1 mM. However, the remaining 42 compounds exhibited average IC<sub>50</sub>s ranging from 29 nM to 979  $\mu$ M. Again the two repeats correlated strongly, Figure 6.52 (a), with a Pearson correlation coefficient of 0.93, computed using GraphPad Prism 7.02<sup>243</sup>, indicative of reproducible data. Apart from the known inhibitors discussed previously, the most potent compounds identified were UOS-32585, L-Moses and UOS-31608, which exhibited IC<sub>50</sub>s of 29 nM, 499 nM and 3.64  $\mu$ M respectively, Figure 6.53. L-Moses, identified as a PCAF BRD inhibitor by Moustakim *et al.* in

2017<sup>581</sup>, was previously tested by ITC and exhibited a  $K_d$  of 2  $\mu\text{M}$  for the KAT2A BRD. UOS-32585 and UOS-31608 are pyridazinone derivatives, similar to UOS-31388.

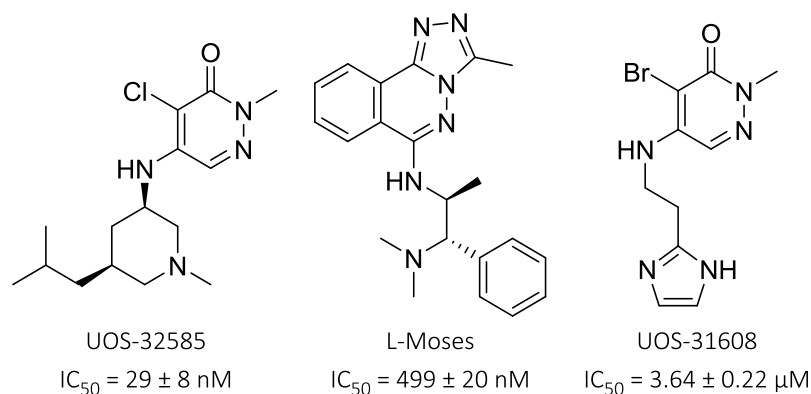


**Figure 6.51. Correlation Between Percentage Inhibition Observed in Two Repeats of Single Point Screens.** 1 mM small molecule added to 0.313 ng anti-GST-Eu<sup>3+</sup>, 80 nM UOS-31670, 10 nM SA-XL665 and 2.5 nM GST-KAT2A BRD, 20  $\mu\text{L}$  well volume, and incubated for two hours. HTRF<sup>®</sup> ratios calculated and background signal, deduced from low controls without GST-KAT2A, subtracted to determine specific signal. Specific signal converted to percentage inhibition dependent on high controls. (a) Screen of 80 small molecules available in-house. (b) Screen of 3140 fragments from Maybridge Ro3 2500 Diversity Fragment Library and Life Chemicals 3D Fragments Library. Graphs compare two repeats.

Following the success of this small compound screen the optimised TR-FRET assay was used to screen the in-house fragment library of 3140 compounds, which is comprised of the Maybridge Ro3 2500 Diversity Fragment Library and the Life Chemicals 3D Fragments Library. The initial single point screen was again conducted in duplicate. 1 mM of each fragment was incubated with 2.5 nM GST-KAT2A, 0.313 ng anti-GST-Eu<sup>3+</sup>, 80 nM UOS-31670 and 10 nM SA-XL665 for two hours. High and low controls were incorporated to determine the maximum and background signals. The fluorescence emission was read at 665 nm and 620 nm wavelengths to attain the HTRF<sup>®</sup> ratios. The specific signals were determined and converted to percentage inhibition using the Studies package from Dotmatics.<sup>244</sup>



**Figure 6.52. Correlation Between  $\text{IC}_{50}$  Values Determined in Two Repeats of Dose-Response Screens.** Ten-point serial dilutions of small molecules diluted 1/3 with top concentration 1 mM, added to 0.313 ng anti-GST-Eu<sup>3+</sup>, 80 nM UOS-31670, 10 nM SA-XL665 and 2.5 nM GST-KAT2A BRD, 20  $\mu\text{L}$  well volume, and incubated for two hours. HTRF<sup>®</sup> ratios calculated and background signal, deduced from low controls without GST-KAT2A, subtracted to determine specific signal. Specific signal converted to percentage inhibition dependent on high controls. (a) Screen of 60 small molecules sourced in-house. (b) Screen of 300 fragments from Maybridge Ro3 2500 Diversity Fragment Library and Life Chemicals 3D Fragments Library. Graphs compare two repeats.



**Figure 6.53. KAT2A BRD Binders Assessed Using TR-FRET Assay.**

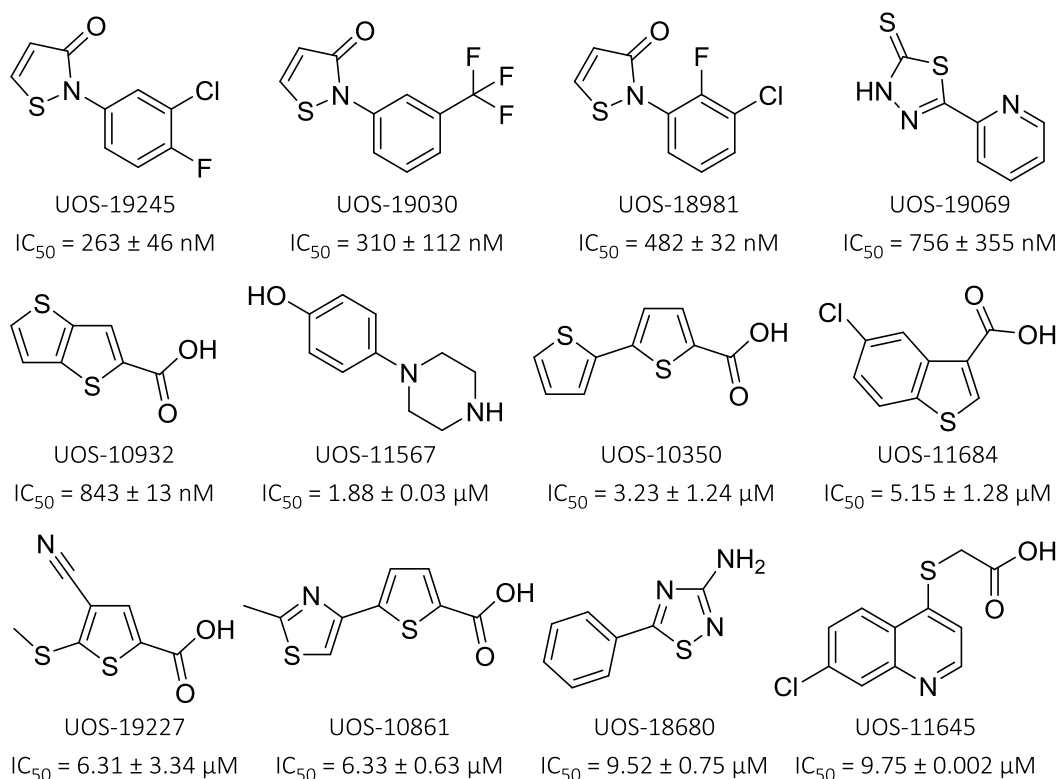
The Vortex package from Dotmatics was utilised for data analysis.<sup>244</sup> The results from the first repeat are displayed in Figure 6.50 (b). As observed in the previous smaller screen, the controls were tightly distributed. In the first and second repeats the positive controls presented standard deviations of just 5.04 and 6.94 respectively. Additionally, the two repeats again correlated strongly, Figure 6.51 (b), with a Pearson correlation coefficient of 0.83, computed using GraphPad Prism 7.02<sup>243</sup>, indicating that the data was reproducible.

As a result of the tight distribution of the controls, inhibition of greater than 18% was deemed significant, calculated as three times the standard deviation in the controls. However, too many fragments inhibited the GST-KAT2A-UOS-31670 binding by more than 18% to progress them all to  $IC_{50}$  analysis. The ten point dilution curves utilised to measure  $IC_{50}$  in this TR-FRET assay employ a top concentration of 1 mM compound. As such,  $IC_{50}$ s greater than 1 mM cannot be determined. Those fragments that inhibit the protein-labelled ligand binding by 50% at 1 mM in the single point screen should exhibit  $IC_{50}$ s of approximately 1 mM. It was therefore decided that any fragments which afforded less than 50% inhibition in the single point screen would be disregarded.

A more manageable set of 300 fragments afforded at least 50% inhibition of GST-KAT2A-UOS-31670 binding in both repeats of the single point screen, equating to 40 dilution plates and 20 assay plates. These fragments were serially diluted at 1/3 to obtain ten point dilution curves with top concentrations of 1 mM. The titrations were incubated with 2.5 nM GST-KAT2A, 0.313 ng anti-GST-Eu<sup>3+</sup>, 80 nM UOS-31670 and 10 nM SA-XL665 for two hours, alongside high and low controls. The fluorescence emission was measured at 665 nm and 620 nm and the HTRF<sup>®</sup> ratios were calculated. The specific signals were calculated and converted into percentage inhibition and the  $IC_{50}$ s determined using the Studies package from Dotmatics.<sup>244</sup>

$IC_{50}$  values could not be determined for 31 of the 300 fragments tested, indicating a value greater than 1 mM. The remaining 269 compounds exhibited average  $IC_{50}$ s ranging from 263 nM to

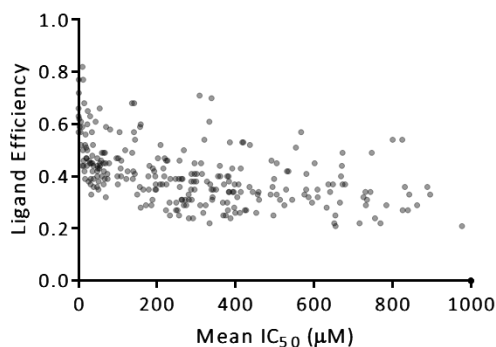
863  $\mu\text{M}$ . Once again the two repeats correlated well, Figure 6.52 (b), with a Pearson correlation coefficient of 0.81, computed using GraphPad Prism 7.02<sup>243</sup>, indicative of good data reproducibility. The 12 most potent fragments are illustrated in Figure 6.54.



**Figure 6.54. KAT2A BRD Fragment Binders Identified Using TR-FRET Assay.**

Unfortunately, the three most potent hits, UOS-19245, UOS-19030 and UOS-18981 were isothiazolones, which are PAINS.<sup>517</sup> As discussed previously in Chapter 5.2, these compounds are frequently identified as artefactual hits due to subversive reactivity that affords false positive signals in a variety of assays, and should therefore be rejected. However, in fragment drug discovery programs it is prudent to also consider ligand efficiency. In addition to the potency, LE takes into account the size of the fragment, determined as the number of non-hydrogen atoms (NHA), and thereby quantifies how effectively a molecule uses its structural features in binding to the target. It is calculated as  $\text{LE} = \Delta\text{G}/\text{NHA} = 1.4(-\log\text{IC}_{50})/\text{NHA}$ . The LEs of the fragments were determined using the Studies package from Dotmatics.<sup>244</sup> Figure 6.55 shows a plot of LE vs. mean  $\text{IC}_{50}$ .

Many of the fragments identified in the screen have good LE values of  $>0.45$  and could make interesting starting points for development in drug discovery projects. It is hoped that a selection of the fragments will represent legitimate and novel hits with affinity for the *Homo sapiens* KAT2A BRD. Unfortunately, due to time constraints, no attempt was made to confirm the hits using any second assay format.



**Figure 6.55. Comparison of Ligand Efficiency and IC<sub>50</sub> Values Determined in Dose-Response Screens.** Mean LEs and IC<sub>50</sub>s calculated from two repeats of ten-point serial dilutions of 300 fragments from Maybridge Ro3 2500 Diversity Fragment Library and Life Chemicals 3D Fragments Library.

### 6.3 Discussion

BRDs represent particularly compelling targets in drug discovery, with many potential disease applications, especially in cancers.<sup>522</sup> They are therefore interesting targets for development of small molecule tool inhibitors, as evidenced by the myriad of BRD inhibitors published in recent years. The focus of this chapter was the KAT2A bromodomain, which was identified in Chapter 4 as an interesting and promising target for tool inhibitor development.

Initially, as at the HAT domain, sequence alignments were used to assess the potential for selectivity. It was again observed that KAT2A is highly homologous with KAT2B, but the residue conservation between the bromodomain families is poorer. This should afford good potential for selectivity except from between KAT2A and KAT2B. A screening cascade was proposed, incorporating T<sub>m</sub> shift, NMR, TR-FRET, ITC and X-ray crystallographic techniques, which are routinely adopted by other research groups in identifying bromodomain tool inhibitors.

A structure-based approach was used to design a library of potential KAT2A BRD binders inspired by the quinolinone scaffold of SL1126, reported as a KAT2B inhibitor by the SGC.<sup>240</sup> Sequence alignments were used to identify residues unique to KAT2A and KAT2B, and libraries of amides and sulphonamides were docked into the PCAF BRD (PDB ID: 5FE8)<sup>240</sup> to assess which might offer improved potency. A T<sub>m</sub> shift assay was established in-house and used to screen this library to identify KAT2A binders. Nine compounds exhibited a significant positive T<sub>m</sub> shift, indicative of an increase in stability of the KAT2A BRD. Only six of these compounds were sufficiently soluble to be investigated using ITC. Of these, two were confirmed as KAT2A binders, including UOS-21430, which exhibited a K<sub>d</sub> of 29 μM.

Simultaneously, an NMR fragment screen was conducted externally to identify novel scaffolds for inhibitor development. 32 fragments were identified that interact with KAT2A. These hit fragments were screened in-house using ITC and seven were confirmed as KAT2A binders.

Due to poor sensitivity and high rate of false positives, the  $T_m$  shift assay was abandoned and ITC was adopted as the primary screen. Three series of quinolinone, phenylpyridinone and pyridylpyridinone derivatives, as well as a collection of fragments were screened. Four compounds were identified which afforded a slight improvement in affinity compared to UOS-21430. The most potent, UOS-31235, exhibited a  $K_d$  of 12  $\mu$ M.

Seven crystal structures were successfully obtained in-house, depicting the KAT2A BRD apo and with small molecules bound. The ligand-bound structures depicted the binding of micromolar KAT2A BRD inhibitors and enhanced understanding of the KAT2A BRD acetyl-lysine binding site. It is hoped that these structures will facilitate improved structure-based drug design.

ITC did not represent an optimal primary screen. It was time consuming, due to low throughput and required a large quantity of protein. As such, a TR-FRET assay was designed and established in-house, which was better suited to high-throughput screening. Attempts to identify a suitable peptide to employ in this competitive binding assay were unproductive. However, at this time, Genentech and Constellation Pharmaceuticals disclosed highly potent dual inhibitors of PCAF and GCN5 in three patent applications, encompassing phthalazine<sup>578</sup>, 6-azauracil<sup>579</sup> and pyridazinone<sup>580</sup> derivatives. Several of the reported inhibitors were tested in-house using ITC, and crystal structures were successfully obtained of two of these, UOS-31388 and UOS-31181, in complex with the KAT2A BRD. UOS-31388, which exhibited a  $K_d$  of 940 nM, was used to develop a ligand-biotin species for use in the TR-FRET competition assay.

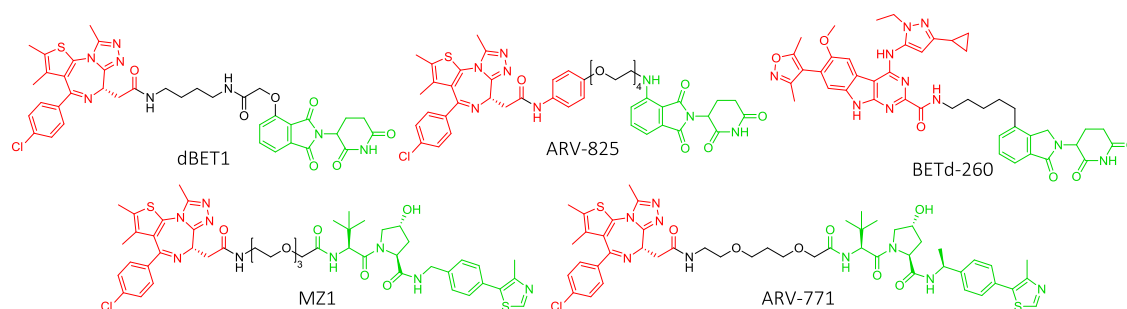
The TR-FRET assay was optimised and used to assess known KAT2A BRD binders. The calculated  $IC_{50}$ s were in good agreement with  $K_d$ s previously determined by ITC. Following this, two single point screens were conducted, the first with 80 compounds available in-house and the second with 3140 fragments from the Maybridge Ro3 2500 Diversity Fragment Library and the Life Chemicals 3D Fragments Library. Hits from the single point screens were progressed to ten-point dose-response screens, used to determine  $IC_{50}$ . The assay proved reliable, the data was reproducible and many novel KAT2A BRD binders were identified. Unfortunately, due to time constraints, it was not possible to confirm these hits using a second assay format, and attempts to develop a KAT2A BRD tool inhibitor were concluded.

Overall, more than 3000 small molecules and fragments were screened in this work, to assess their affinity for KAT2A, and many novel KAT2A BRD inhibitors were discovered. The optimised assays, novel chemical matter and ligand-bound crystal structures afford an exciting opportunity for the development of potent and selective KAT2A BRD tool inhibitors.



As discussed, bromodomains are protein-protein interaction domains, which recognise and bind to acetylated lysine residues and thereby direct protein function.<sup>522</sup> It is hoped that KAT2A BRD inhibitors will disrupt interactions with the acetylated protein targets and thereby disrupt KAT2A function by preventing the protein from properly localising to target sites. However, BRD inhibitors will not impede the catalytic activity of KAT2A and therefore may not afford the desired biological response and selectively kill BAF180 deficient cells. To navigate this obstacle it is proposed that proteolysis targeting chimera (PROTAC) technology could be employed.<sup>583</sup>

A number of PROTACs have been developed, which target BET bromodomain proteins for degradation, Figure 6.56. These chimeric molecules comprise of BET inhibitors tethered to E3 ubiquitin ligase modulators via a flexible linker. Upon binding the PROTAC, the BET BRD protein is selectively ubiquitinated and degraded via a proteasome-mediated mechanism.<sup>545</sup> As such, it is possible to inhibit the catalytic activity of the target BRD protein using compounds which bind at the bromodomain. PROTACs MZ1<sup>584</sup> and ARV-771<sup>585</sup> were designed from BET inhibitor JQ1<sup>217</sup> tethered to VHL ligase modulators. dBET1<sup>586</sup> and ARV-825<sup>587</sup> were derived from BET inhibitors JQ1<sup>217</sup> and OTX015<sup>554</sup> coupled with thalidomide and pomalidomide respectively, which are modulators of cereblon, a component of the cullin-RING ubiquitin ligase complex. Finally, BETd-260 was designed from an azocarbazole-based BET inhibitor<sup>588</sup>, with lenalidomide, another cereblon modulator.<sup>589</sup> These BET degraders have been demonstrated to efficiently induce BET protein degradation and are more potent in inhibiting cancer cell growth and inducing apoptosis than the associated BET inhibitors.<sup>589</sup> Small molecule BRD protein degraders therefore represent promising therapeutics for treatment of cancers and other human diseases, and KAT2A protein degraders may enable selective killing of BAF180 deficient cells even if bromodomain tool inhibitors do not.



**Figure 6.56. Chemical Structures of Reported BET BRD PROTACs.** Chimeric molecules composed of BET inhibitors (red) conjugated to E3 ubiquitin ligase modulators (green) via flexible linkers, which induce proteasome-mediated degradation of BET proteins.

## 7 Discussion

As a result of the high cost of drug discovery and poor productivity of research and development<sup>10</sup>, it is imperative that more effort is expended to identify and validate the best targets with the strongest disease association before embarking on the costly molecule discovery and development phases. Good quality tool inhibitors provide an excellent mechanism to test therapeutic hypotheses. A non-toxic, therapeutic phenotype in cellular or *in vivo* models in response to a probe is fundamental in target validation.<sup>216</sup> Additionally, chemical probes are remarkably powerful in unveiling and accessing novel biology and potential targets.<sup>214</sup> Consequently, publication of a new tool inhibitor often provokes increased research in the area.

This project was initiated with the aim of designing and characterising tool inhibitors of proteins involved in DNA damage response. The areas of primary focus were the structural maintenance of chromosomes proteins and lysine acetyltransferase 2-A. It was hoped that such probes would enable detailed mechanistic biological investigation, disease validation and ultimately initiation of translational drug discovery projects.

### 7.1 SMC Complexes

The SMC complexes are critically important in regulating chromosome condensation, sister chromatid cohesion, DNA repair, homologous recombination and transcription. They represent interesting targets for tool inhibitor development, particularly as these roles in regulating chromosome architecture and organisation are well known, but not well understood.

In this work, the initial focus was SMC5/6, which has a prominent role in DNA repair<sup>254–256</sup> and the ALT pathway<sup>361</sup>, and represents a potential opportunity for novel cancer therapeutics dependent on synthetic lethality. The SMC5/6 hinge domain and head domain and the non-SMC elements were modelled and analysed to identify whether they represent druggable targets, suited to tool inhibitor development.

Unfortunately, the SMC5/6 hinge domain was deemed undruggable as there were no discernible structural pockets and therefore no potential for rational inhibitor design. Potential binding pockets were identified at the SMC5 N-terminal ATP binding site. Although the druggability scores<sup>170</sup> suggested that these pockets were not druggable, it was reasoned that this could result from the significant solvent exposure or the polar environment at the site. However, in ITC binding studies, the SMC5 head domain was not observed to interact with ATP derivatives, indicating that the pocket is not suitable for small molecule binding, and therefore the SMC5/6 head domain was also deemed undruggable. In the models of the NSEs, cavities were identified

at the NSE1-NSE3 and NSE2-SMC5 PPI interfaces, but neither represented desirable target sites for tool inhibitor development.

Following this, the human cohesin hinge and head domains were modelled and analysed. A pocket was identified in the SMC3 subunit at the hinge domain, which corresponded to a helix in SMC1, but the fpocket druggability score<sup>170</sup> indicated that this cavity was not druggable.

In conclusion, at the time this work was completed, the SMC5/6 and cohesin complexes did not represent desirable targets for design and characterisation of tool inhibitors.

## 7.2 BAF180 Synthetic Lethal Targets

BAF180 has been identified as a major clear cell renal cell carcinoma cancer gene, with truncating mutations observed in 41% of samples of primary ccRCC<sup>1</sup>, and has therefore been highlighted as a promising opportunity to target ccRCC using a synthetic lethality approach.<sup>2</sup> RNF4, ASF1A and GCN5/KAT2A have been identified as synthetic lethal partners of PBRM1<sup>3</sup>, and therefore may represent interesting targets for tool inhibitor development.

In this work, RNF4, ASF1A and GCN5 were analysed to identify any potentially druggable binding pockets. *Rattus norvegicus* RNF4 (PDB ID: 3NG2)<sup>234</sup> exhibited one interesting druggable pocket, but no molecules could be identified which bound at this site to indicate any corresponding pharmacophore, so it was concluded that RNF4 did not represent a promising target. *Homo sapiens* ASF1A (PDB ID: 2IO5)<sup>236</sup> similarly exhibited one interesting druggable pocket, but this was located at the ASF1A-HIRA interface. In unpublished work (Hopkins S. and Downs J., Genome Damage and Stability Centre, University of Sussex), it has been shown that HIRA and BAF180 are not synthetic lethal. It was therefore concluded that ASF1A also did not represent a good target.

KAT2A contains a HAT domain and a bromodomain. Both *Homo sapiens* GCN5 Acetyltransferase domain (PDB ID: 1Z4R)<sup>241</sup> and *Homo sapiens* GCN5 bromodomain, (PDB ID: 3D7C)<sup>238</sup>, exhibited druggable pockets, which coincided with the acetyl-CoA and acetyl-lysine binding sites respectively. KAT2A was therefore established as a promising and interesting target for tool inhibitor development.

## 7.3 KAT2A HAT Domain

Based on these results, KAT2A became the focus of this project. With druggable pockets identified in both the HAT and BRD structural domains, the catalytically active HAT domain was prioritised as the preferred target. HATs currently represent exciting targets in drug discovery,

with many potential disease applications, and are therefore interesting targets for tool inhibitor development, especially as existing HAT inhibitors are not suitable for biological investigation.

In this work, sequence alignments were used to assess the potential for selectivity between KAT2A and other HAT-containing proteins. KAT2A shows extremely high homology with KAT2B, but overall the residue conservation throughout HATs is poor, which should afford good selectivity except from between KAT2A and KAT2B.

Attempts to establish a fluorescence based activity assay in-house, which had previously been used in HTS for KAT2A HAT inhibitors (PubChem BioAssay AID: 504327)<sup>4</sup>, were unsuccessful. The assay proved unreliable and robust activity could not be achieved. Initial attempts to establish an alternative EpiQuik™ HAT activity assay were promising but associated with high error.

The hits from the HTS<sup>4,5</sup> were discounted as likely false positives, which significantly diminished the inventory of reported small molecule inhibitors of KAT2A and KAT2B. The only remaining series was the isothiazolones, which are well known PAINS and were therefore also disregarded.

Unfortunately, while HATs represent interesting targets for tool inhibitor development, in general they have proved very challenging to work on. However, novel HAT inhibitors remain necessary and highly desirable. It is vital that an alternative KAT2A HAT assay format is developed, such as radioligand binding, which offers robust activity, and that a high-throughput screen is conducted to identify new reliable chemical matter. This was regrettably beyond the scope of this work and attempts to develop a KAT2A HAT tool inhibitor were concluded.

#### 7.4 KAT2A Bromodomain

Despite the difficulties encountered in developing a tool inhibitor for the HAT domain, KAT2A remained an interesting protein target. As such, focus was redirected to the KAT2A BRD, which was also identified as a druggable target in Chapter 4.2. Bromodomains are a hot topic in research, with many disease applications, particularly in cancers<sup>522</sup>, and represent particularly compelling targets for tool inhibitor development.

Initially, sequence alignments were used to assess the potential for selectivity between KAT2A and other BRD-containing proteins. KAT2A is highly homologous with KAT2B, but the conservation between BRD families is poor. This should afford good selectivity in general, but again it is anticipated that inhibitors will target both KAT2A and KAT2B.

A structure-based design approach was used to generate a library of potential KAT2A BRD binders, utilising the ligand-bound structure of quinolinone SL1126 with the PCAF BRD, (PDB ID:

5FE8).<sup>240</sup> The library of compounds was screened in-house using T<sub>m</sub> shift and ITC. Two compounds were confirmed as KAT2A binders, including UOS-21430, with a K<sub>d</sub> of 29 μM.

An NMR fragment screen was conducted externally to identify novel scaffolds for inhibitor development. The 32 hit fragments were screened in-house using ITC and seven were confirmed as KAT2A binders.

The T<sub>m</sub> shift assay was abandoned due to poor sensitivity and high rate of false positives and ITC was adopted as the primary screen to assess three series of quinolinone, phenylpyridinone and pyridylpyridinone derivatives, as well as a number of fragments. Four of the compounds afforded a slight improvement in affinity compared to UOS-21430. The most potent was UOS-31235, with a K<sub>d</sub> of 12 μM.

ITC did not represent an optimal primary screen. It was time consuming, low throughput and required a large supply of protein. As such, a TR-FRET assay was designed and established in-house, which was better suited to high-throughput screening. Typically, these competitive binding assays employ biotin-labelled peptides, however, in this case, a ligand-biotin species was utilised, inspired by the pyridazinone derivatives disclosed as potent dual inhibitors of PCAF and GCN5 in the patent application from Genentech and Constellation Pharmaceuticals.<sup>580</sup>

The TR-FRET assay was optimised and used to assess known KAT2A BRD binders. The calculated IC<sub>50</sub>s were in good agreement with K<sub>d</sub>s determined by ITC. Subsequently, two single point screens were conducted, the first with 80 compounds and the second with 3140 fragments. Hits from the single point screens were progressed to ten-point dose-response screens, which were used to determine IC<sub>50</sub>s. The assay format was reliable, the data was reproducible and many novel KAT2A BRD binders were identified. Unfortunately, it was not possible to confirm the hits using a second assay format, due to time constraints.

Alongside screening, nine crystal structures were successfully obtained in-house, depicting the apo and ligand-bound KAT2A BRD. It is hoped that these structures will facilitate improved structure-based drug design.

Overall, in this work, more than 3000 small molecules and fragments were screened using a combination of assay formats, and many novel KAT2A BRD inhibitors were discovered. The optimised assays, novel chemical matter and ligand-bound crystal structures afford an exciting opportunity to develop potent and selective KAT2A BRD tool inhibitors.

## VI References

1. Varela, I. *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* **469**, 539–542 (2011).
2. Chan, D. A. & Giaccia, A. J. Harnessing synthetic lethal interactions in anticancer drug discovery. *Nat. Rev. Drug Discov.* **10**, 351–364 (2011).
3. Hopkins, S. R., McGregor, G. A., Murray, J. M., Downs, J. A. & Savic, V. Novel synthetic lethality screening method identifies TIP60-dependent radiation sensitivity in the absence of BAF180. *DNA Repair (Amst)*. **46**, 47–54 (2016).
4. National Center for Biotechnology Information. PubChem BioAssay Database; AID=504327, Source=NIH Chemical Genomics Center. at <<https://pubchem.ncbi.nlm.nih.gov/bioassay/504327>>
5. National Center for Biotechnology Information. PubChem BioAssay Database; AID=588347, Source=NIH Chemical Genomics Center. at <<https://pubchem.ncbi.nlm.nih.gov/bioassay/588347>>
6. Dimasi, J. A., Grabowski, H. G. & Hansen, R. W. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J. Health Econ.* **47**, 20–33 (2016).
7. Paul, S. M. *et al.* How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* **9**, 203–214 (2010).
8. Hughes, J. P., Rees, S., Kalindjian, S. B. & Philpott, K. L. Principles of early drug discovery. *Br. J. Pharmacol.* **162**, 1239–1249 (2011).
9. Tari, L. W. *The Utility of Structural Biology in Drug Discovery. Structure-Based Drug Discovery, Methods in Molecular Biology, vol. 841* (Humana Press, 2012). doi:10.1007/978-1-61779-520-6
10. Scannell, J. W., Blanckley, A., Boldon, H. & Warrington, B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov.* **11**, 191–200 (2012).
11. Schnee, J. E. R&D strategy in the U.S. pharmaceutical industry. *Res. Policy* **8**, 364–382 (1979).
12. Williams, M. Productivity Shortfalls in Drug Discovery: Contributions from the Preclinical Sciences? *Pharmacol. Exp. Ther.* **336**, 3–8 (2011).
13. Waring, M. J. *et al.* An analysis of the attrition of pharmaceutical companies. *Nat. Rev. Drug Discov.* **14**, 475–486 (2015).
14. Plenge, R. M. Disciplined approach to drug discovery and early development. *Sci. Transl. Med.* **8**, 1–6 (2016).
15. Lindahl, T. & Barnes, D. E. Repair of endogenous DNA damage. *Cold Spring Harb. Symp. Quant. Biol.* **65**, 127–133 (2000).
16. Jackson, S. P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* **461**, 1071–1078 (2009).
17. Rouse, J. & Jackson, S. P. Interfaces between the detection, signaling, and repair of DNA damage. *Science (80-. )*. **297**, 547–551 (2002).
18. Maréchal, A. & Zou, L. DNA damage sensing by the ATM and ATR kinases. *Cold Spring Harb. Perspect. Biol.* **5**, 1–17 (2013).
19. Lempiäinen, H. & Halazonetis, T. D. Emerging common themes in regulation of PIKKs

- and PI3Ks. *EMBO J.* **28**, 3067–3073 (2009).
20. Sirbu, B. M. & Cortez, D. DNA damage response: three levels of DNA repair regulation. *Cold Spring Harb. Perspect. Biol.* **5**, 1–16 (2013).
  21. Falck, J., Coates, J. & Jackson, S. P. Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. *Nature* **434**, 605–611 (2005).
  22. Perry, J. & Kleckner, N. The ATRs, ATMs, and TORs are Giant HEAT Repeat Proteins. *Cell* **112**, 151–155 (2003).
  23. Sibanda, B. L., Chirgadze, D. Y. & Blundell, T. L. Crystal Structure of DNA-PKcs Reveals a Large Open-Ring Cradle Comprised of HEAT Repeats. *Nature* **463**, 118–121 (2010).
  24. You, Z., Chahwan, C., Bailis, J., Hunter, T. & Russell, P. ATM activation and its recruitment to damaged DNA require binding to the C terminus of Nbs1. *Mol. Cell. Biol.* **25**, 5363–5379 (2005).
  25. Ball, H. L., Myers, J. S. & Cortez, D. ATRIP Binding to Replication Protein A-Single-stranded DNA Promotes ATR-ATRIP Localization but Is Dispensable for Chk1 Phosphorylation. *Mol. Biol. Cell* **16**, 2372–2381 (2005).
  26. Bakkenist, C. J. & Kastan, M. B. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature* **421**, 499–506 (2003).
  27. Liu, S. *et al.* ATR Autophosphorylation as a Molecular Switch for Checkpoint Activation. *Mol. Cell* **43**, 192–202 (2011).
  28. Sun, Y., Jiang, X., Chen, S., Fernandes, N. & Price, B. D. A role for the Tip60 histone acetyltransferase in the acetylation and activation of ATM. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 13182–13187 (2005).
  29. Nickoloff, J. A. *DNA Repair Dysregulation in Cancer: From Molecular Mechanisms to Synthetic Lethal Opportunities. Stress Response Pathways in Cancer: From Molecular Targets to Novel Therapeutics* (Springer, 2015). doi:10.1007/978-94-017-9421-3
  30. Lee, J. H. & Paull, T. T. ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. *Science (80-. ).* **308**, 551–554 (2005).
  31. Lavin, M., Kozlov, S., Gatei, M. & Kijas, A. ATM-Dependent Phosphorylation of All Three Members of the MRN Complex: From Sensor to Adaptor. *Biomolecules* **5**, 2877–2902 (2015).
  32. Gottlieb, T. M. & Jackson, S. P. The Dna-Dependent Protein-Kinase: Requirement for Dna Ends and Association With Ku Antigen. *Cell* **72**, 131–142 (1993).
  33. Matsuoka, S. *et al.* ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science (80-. ).* **316**, 1160–1166 (2007).
  34. Zou, L. & Elledge, S. J. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science (80-. ).* **300**, 1542–1548 (2003).
  35. Acevedo, J., Yan, S. & Michael, W. M. Direct binding to replication protein A (RPA)-coated single-stranded DNA allows recruitment of the ATR activator TopBP1 to sites of DNA damage. *J. Biol. Chem.* **291**, 13124–13131 (2016).
  36. Zou, L., Cortez, D. & Elledge, S. J. Regulation of ATR substrate selection by Rad17-dependent loading of Rad9 complexes onto chromatin. *Genes Dev.* **16**, 198–208 (2002).
  37. Kumagai, A., Lee, J., Yoo, H. Y. & Dunphy, W. G. TopBP1 activates the ATR-ATRIP complex. *Cell* **124**, 943–955 (2006).

38. Mordes, D. A., Glick, G. G., Zhao, R. & Cortez, D. TopBP1 activates ATR through ATRIP and a PIKK regulatory domain. *Genes Dev.* **22**, 1478–1489 (2008).
39. Shiotani, B. *et al.* Two Distinct Modes of ATR Activation Orchestrated by Rad17 and Nbs1. *Cell Rep.* **3**, 1651–1662 (2013).
40. Deans, A. J. & West, S. C. DNA interstrand crosslink repair and cancer. *Nat. Rev. Cancer* **11**, 467–480 (2011).
41. Swuec, P. & Costa, A. DNA replication and inter-strand crosslink repair: Symmetric activation of dimeric nanomachines? *Biophys. Chem.* 1–5 (2016). doi:10.1016/j.bpc.2016.11.001
42. Ishiai, M. *et al.* FANCI phosphorylation functions as a molecular switch to turn on the Fanconi anemia pathway. *Nat. Struct. Mol. Biol.* **15**, 1138–1146 (2008).
43. Fernandez-Capetillo, O., Lee, A., Nussenzweig, M. & Nussenzweig, A. H2AX: The histone guardian of the genome. *DNA Repair (Amst)*. **3**, 959–967 (2004).
44. Al-Hakim, A. *et al.* The ubiquitous role of ubiquitin in the DNA damage response. *DNA Repair (Amst)*. **9**, 1229–1240 (2010).
45. Panier, S. & Boulton, S. J. Double-strand break repair: 53BP1 comes into focus. *Nat. Rev. Mol. Cell Biol.* **15**, 7–18 (2014).
46. Lee, H.-S., Park, J.-H., Kim, S.-J., Kwon, S.-J. & Kwon, J. A cooperative activation loop among SWI/SNF, gamma-H2AX and H3 acetylation for DNA double-strand break repair. *EMBO J.* **29**, 1434–1445 (2010).
47. Peng, G. *et al.* BRIT1/MCPH1 links chromatin remodelling to DNA damage response. *Nat. Cell Biol.* **11**, 865–872 (2009).
48. Da-Rè, C. & Halazonetis, T. D. DNA replication stress as an Achilles' heel of cancer. *Oncotarget* **6**, 1–2 (2015).
49. Krokan, H. E. & Bjørås, M. Base excision repair. *Cold Spring Harb. Perspect. Biol.* **5**, 1–22 (2013).
50. Prasad, R., Dyrkheeva, N., Williams, J. & Wilson, S. H. Mammalian base excision repair: Functional partnership between PARP-1 and APE1 in AP-site repair. *PLoS One* **10**, 1–19 (2015).
51. Hegde, M. L., Hazra, T. K. & Mitra, S. Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells. *Cell Res.* **18**, 27–47 (2008).
52. Kamileri, I., Karakasilioti, I. & Garinis, G. A. Nucleotide excision repair: New tricks with old bricks. *Trends Genet.* **28**, 566–573 (2012).
53. Hsieh, P. & Zhang, Y. The Devil is in the details for DNA mismatch repair. *Proc. Natl. Acad. Sci.* **114**, 3552–3554 (2017).
54. Simonelli, V., Narciso, L., Dogliotti, E. & Fortini, P. Base excision repair intermediates are mutagenic in mammalian cells. *Nucleic Acids Res.* **33**, 4404–4411 (2005).
55. Deriano, L. & Roth, D. B. Modernizing the Nonhomologous End-Joining Repertoire: Alternative and Classical NHEJ Share the Stage. *Annu. Rev. Genet.* **47**, 433–55 (2013).
56. Jacobs, C. *et al.* A hypomorphic Artemis human disease allele causes aberrant chromosomal rearrangements and tumorigenesis. *Hum. Mol. Genet.* **20**, 806–819 (2011).
57. Fnu, S. *et al.* Methylation of histone H3 lysine 36 enhances DNA repair by



- nonhomologous end-joining. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 540–545 (2011).
58. Sharma, S. *et al.* Homology and enzymatic requirements of microhomology-dependent alternative end joining. *Cell Death Dis.* **6**, e1697 (2015).
  59. Symington, L. S. & Gautier, J. Double-Strand Break End Resection and Repair Pathway Choice. *Annu. Rev. Genet.* **45**, 247–271 (2011).
  60. Heyer, W.-D., Ehmsen, K. T. & Liu, J. Regulation of homologous recombination in eukaryotes. *Annu. Rev. Genet.* **44**, 113–139 (2010).
  61. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).
  62. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).
  63. Witsch, E., Sela, M. & Yarden, Y. Roles for Growth Factors in Cancer Progression. *Physiology* **25**, 85–101 (2010).
  64. Bhowmick, N. A., Neilson, E. G. & Moses, H. L. Stromal fibroblasts in cancer initiation and progression. *Nature* **432**, 332–337 (2004).
  65. Davies, M. A. & Samuels, Y. Analysis of the genome to personalize therapy for melanoma. *Oncogene* **29**, 5545–5555 (2010).
  66. Yuan, T. L. & Cantley, L. C. PI3K pathway alterations in cancer: variations on a theme. *Oncogene* **27**, 5497–5510 (2008).
  67. Sherr, C. J. & McCormick, F. The RB and p53 pathways in cancer. *Cancer Cell* **2**, 103–112 (2002).
  68. Curto, M., Cole, B. K., Lallemand, D., Liu, C. H. & McClatchey, A. I. Contact-dependent inhibition of EGFR signaling by Nf2/Merlin. *J. Cell Biol.* **177**, 893–903 (2007).
  69. Lowe, S. W., Cepero, E. & Evan, G. Intrinsic tumour suppression. *Nature* **432**, 307–315 (2004).
  70. Adams, J. M. & Cory, S. The Bcl-2 apoptotic switch in cancer development and therapy. *Oncogene* **26**, 1324–37 (2007).
  71. Junttila, M. R. & Evan, G. I. p53 - a Jack of all trades but master of none. *Nat. Rev. Cancer* **9**, 821–829 (2009).
  72. Grivennikov, S. I., Greten, F. R. & Karin, M. Immunity, Inflammation, and Cancer. *Cell* **140**, 883–899 (2010).
  73. Blasco, M. A. Telomeres and human disease: ageing, cancer and beyond. *Nat. Reviews. Genet.* **6**, 611–622 (2005).
  74. Artandi, S. E. & DePinho, R. A. Telomeres and telomerase in cancer. *Carcinogenesis* **31**, 9–18 (2009).
  75. Hanahan, D. & Folkman, J. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell* **86**, 353–364 (1996).
  76. Baeriswyl, V. & Christofori, G. The angiogenic switch in carcinogenesis. *Semin. Cancer Biol.* **19**, 329–337 (2009).
  77. Kazerounian, S., Yee, K. O. & Lawler, J. Thrombospondins in cancer. *Cell. Mol. Life Sci.* **65**, 700–712 (2008).
  78. Zumsteg, A. & Christofori, G. Corrupt policemen: inflammatory cells promote tumor angiogenesis. *Curr. Opin. Oncol.* **21**, 60–70 (2009).

79. Jin, K., Li, T., van Dam, H., Zhou, F. & Zhang, L. Molecular insights into tumour metastasis: tracing the dominant events. *J. Pathol.* **241**, 567–577 (2017).
80. Massagué, J. TGF $\beta$  in Cancer. *Cell* **134**, 215–230 (2008).
81. Micalizzi, D. S., Farabaugh, S. M. & Ford, H. L. Epithelial-mesenchymal transition in cancer: Parallels between normal development and tumor progression. *J. Mammary Gland Biol. Neoplasia* **15**, 117–134 (2010).
82. Brabletz, T. *et al.* Variable beta-catenin expression in colorectal cancers indicates tumor progression driven by the tumor environment. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10356–10361 (2001).
83. Joyce, J. A. & Pollard, J. W. Microenvironmental regulation of metastasis. *Nat. Rev. Cancer* **9**, 239–252 (2009).
84. Warburg, O. On the Origin of Cancer Cells. *Science (80- )*. **123**, 309–314 (1956).
85. Potter, V. R. The biochemical approach to the cancer problem. *Fed. Proc.* **17**, 691–697 (1958).
86. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg Effect: Cell Proliferation. *Science (80- )*. **324**, 1029–1034 (2009).
87. DeBerardinis, R. J., Lum, J. J., Hatzivassiliou, G. & Thompson, C. B. The Biology of Cancer: Metabolic Reprogramming Fuels Cell Growth and Proliferation. *Cell Metab.* **7**, 11–20 (2008).
88. Kim, R., Emi, M. & Tanabe, K. Cancer immunoediting from immune surveillance to immune escape. *Immunology* **121**, 1–14 (2007).
89. Bindea, G., Mlecnik, B., Fridman, W. H., Pagès, F. & Galon, J. Natural immunity to cancer in humans. *Curr. Opin. Immunol.* **22**, 215–222 (2010).
90. Yang, L., Pang, Y. & Moses, H. L. TGF- $\beta$  and immune cells: an important regulatory axis in the tumor microenvironment and progression. *Trends Immunol.* **31**, 220–227 (2010).
91. Khalil, D. N., Smith, E. L., Brentjens, R. J. & Wolchok, J. D. The future of cancer treatment: Immunomodulation, CARs and combination immunotherapy. *Nat. Rev. Clin. Oncol.* **13**, 273–290 (2016).
92. Baer, C. F., Miyamoto, M. M. & Denver, D. R. Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nat. Rev. Genet.* **8**, 619–631 (2007).
93. von Hanseemann, D. P. Über asymmetrische Zelltheilung in Epithelkrebsen und deren biologische Bedeutung. *Virchow's Arch. für Pathol. Anat. und Physiol. und für Klin. Med.* **779**, 299–326 (1890).
94. Boveri, T. Zur Frage der Entstehung maligner Tumoren. *Fischer Jena* 1–64 (1914).
95. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396**, 643–649 (1998).
96. Shih, I. *et al.* Evidence That Genetic Instability Occurs at an Early Stage of Colorectal Tumorigenesis Advances in Brief Evidence That Genetic Instability Occurs at an Early Stage of Tumorigenesis. *Cancer Res.* **61**, 818–822 (2001).
97. Hanks, S. *et al.* Constitutional aneuploidy and cancer predisposition caused by biallelic mutations in BUB1B. *Nat. Genet.* **36**, 1159–1161 (2004).
98. Fox, E. J. & Loeb, L. A. Lethal Mutagenesis: Targeting the Mutator Phenotype in Cancer. *Semin. Cancer Biol.* **20**, 353–359 (2010).

99. Rubin, A. F. & Green, P. Mutation patterns in cancer genomes. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 21766–21770 (2009).
100. Kaelin, W. G. The Concept of Synthetic Lethality in the Context of Anticancer Therapy. *Nat. Rev. Cancer* **5**, 689–698 (2005).
101. Allen, C., Ashley, A. K., Hromas, R. & Nickoloff, J. A. More forks on the road to replication stress recovery. *J. Mol. Cell Biol.* **3**, 4–12 (2011).
102. Gururangan, S. Late effects of chemotherapy. *Cancer Treat. Res.* **150**, 43–65 (2009).
103. Reddy, A. & Kaelin Jr., W. G. Using Cancer Genetics to Guide the Selection of Anticancer Drug Targets. *Curr Opin Pharmacol* **2**, 366–373 (2002).
104. Liu, Y., Sun, J. & Zhao, M. ONGene: A Literature-Based Database For Human Oncogenes. *J. Genet. Genomics* 10–12 (2016). doi:10.1016/j.jgg.2016.12.004
105. Pasic, I. & Lipton, J. H. Current Approach to the Treatment of Chronic Myeloid Leukaemia. *Leuk. Res.* **55**, 65–78 (2017).
106. Dobzhansky, T. H. Genetics of Natural Populations. XIII. Recombination and Variability in Populations of *Drosophila Pseudoobscura*. *Genetics* **31**, 269–290 (1946).
107. Hartwell, L. H., Szankasi, P., Roberts, C. J., Murray, A. W. & Friend, S. H. Integrating Genetic Approaches into the Discovery of Anticancer Drugs. *Science (80-. )*. **278**, 1064–1068 (1997).
108. Nijman, S. M. B. Synthetic lethality: General principles, utility and detection using genetic screens in human cells. *FEBS Lett.* **585**, 1–6 (2011).
109. Bryant, H. E. *et al.* Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**, 913–917 (2005).
110. Farmer, H. *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917–921 (2005).
111. Wooster, R. & Weber, B. L. Breast and ovarian cancer. *N. Engl. J. Med.* **348**, 2339–2347 (2003).
112. Schultz, N., Lopez, E., Saleh-Gohari, N. & Helleday, T. Poly(ADP-ribose) polymerase (PARP-1) has a controlling role in homologous recombination. *Nucleic Acids Res.* **31**, 4959–4964 (2003).
113. Garber, K. PARP inhibitors bounce back. *Nat. Rev. Drug Discov.* **12**, 725–727 (2013).
114. Settleman, J. & Cohen, R. L. Communication in Drug Development: ‘Translating’ Scientific Discovery. *Cell* **164**, 1101–1104 (2016).
115. Wang, Y. Q. *et al.* An update on poly(ADP-ribose)polymerase-1 (PARP-1) inhibitors: Opportunities and challenges in cancer therapy. *J. Med. Chem.* **59**, 9575–9598 (2016).
116. Yuan, B. *et al.* Poly(ADP-ribose)polymerase (PARP) inhibition and anticancer activity of simmiparib, a new inhibitor undergoing clinical trials. *Cancer Lett.* **386**, 47–56 (2017).
117. Lord, C. J. & Ashworth, A. Mechanisms of resistance to therapies targeting BRCA-mutant cancers. *Nat. Med.* **19**, 1381–1388 (2013).
118. Pang, X. & Liu, M. Defeat Mutant KRAS with Synthetic Lethality. *Small GTPases* **0**, 1–8 (2016).
119. Bolognesi, M. L. & Cavalli, A. Multitarget Drug Discovery and Polypharmacology. *ChemMedChem* **11**, 1190–1192 (2016).

120. Simon, J. A. *et al.* Differential Toxicities of Anticancer Agents among DNA Repair and Checkpoint Mutants of *Saccharomyces cerevisiae*. *Cancer Res.* **60**, 328–333 (2000).
121. Torrance, C. J., Agrawal, V., Vogelstein, B. & Kinzler, K. W. Use of isogenic human cancer cells for high-throughput screening and drug discovery. *Nat. Biotechnol.* **19**, 940–945 (2001).
122. Shaw, A. T. *et al.* Selective killing of K-ras mutant cancer cells by small molecule inducers of oxidative stress. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 8773–8778 (2011).
123. Schenone, M., Dančík, V., Wagner, B. K. & Clemons, P. A. Target identification and mechanism of action in chemical biology and drug discovery. *Nat. Chem. Biol.* **9**, 232–240 (2013).
124. Kuai, L. *et al.* AAK1 identified as an inhibitor of neuregulin-1/ErbB4-dependent neurotrophic factor signaling using integrative chemical genomics and proteomics. *Chem. Biol.* **18**, 891–906 (2011).
125. Edgar, K. A. *et al.* Synthetic lethality of retinoblastoma mutant cells in the *Drosophila* eye by mutation of a novel peptidyl prolyl isomerase gene. *Genetics* **170**, 161–171 (2005).
126. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
127. Cerutti, H. & Casas-Mollano, J. A. On the origin and functions of RNA-mediated silencing: From protists to man. *Curr. Genet.* **50**, 81–99 (2006).
128. Fraser, A. G. *et al.* Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408**, 325–330 (2000).
129. Gönczy, P. *et al.* Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408**, 331–336 (2000).
130. Willingham, A. T., Deveraux, Q. L., Hampton, G. M. & Aza-Blanc, P. RNAi and HTS: exploring cancer by systematic loss-of-function. *Oncogene* **23**, 8392–8400 (2004).
131. Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).
132. Rana, T. M. Illuminating the silence: understanding the structure and function of small RNAs. *Nat. Rev. Mol. Cell Biol.* **8**, 23–36 (2007).
133. Alberts, B. *et al.* *Molecular Biology of the Cell*. (Garland Science, 2008).
134. Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. & Hannon, G. J. Argonaute2, a Link Between Genetic and Biochemical Analyses of RNAi. *Science* (80-. ). **293**, 1146–1150 (2001).
135. Kim, V. N. MicroRNA biogenesis: coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.* **6**, 376–385 (2005).
136. Bohnsack, M. T., Czaplinski, K. & Görlich, D. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* **10**, 185–191 (2004).
137. De Paula, D., Bentley, M. V. L. B. & Mahato, R. I. Hydrophobization and bioconjugation for enhanced siRNA delivery and targeting. *RNA* **13**, 431–456 (2007).
138. Iorns, E., Lord, C. J., Turner, N. & Ashworth, A. Utilizing RNA interference to enhance cancer drug discovery. *Nat. Rev. Drug Discov.* **6**, 556–568 (2007).
139. Mohr, S. E., Smith, J. A., Shamu, C. E., Neumüller, R. A. & Perrimon, N. RNAi screening

- comes of age: improved techniques and complementary approaches. *Nat. Reviews. Mol. Cell Biol.* **15**, 591–600 (2014).
140. Hu, Y. *et al.* UP-TORR: Online tool for accurate and up-to-date annotation of RNAi reagents. *Genetics* **195**, 37–45 (2013).
  141. Brummelkamp, T. R. *et al.* An shRNA barcode screen provides insight into cancer cell vulnerability to MDM2 inhibitors. *Nat. Chem. Biol.* **2**, 202–206 (2006).
  142. Echeverri, C. J. *et al.* Minimizing the risk of reporting false positives in large-scale RNAi screens. *Nat. Methods* **3**, 777–779 (2006).
  143. Kaelin, W. G. Use and Abuse of RNAi to Study Mammalian Gene Function. *Science* (80-.). **337**, 421–422 (2012).
  144. Bhinder, B. & Djaballah, H. Systematic analysis of RNAi reports identifies dismal commonality at gene-level & reveals an unprecedented enrichment in pooled shRNA screens. *Comb. Chem. High Throughput Screen.* **16**, 665–681 (2013).
  145. Tong, A. H. Y. & Boone, C. Synthetic genetic array analysis in *Saccharomyces cerevisiae*. *Methods Molecular Biol.* **313**, 171–191 (2006).
  146. Guénolé, A. *et al.* Dissection of DNA Damage Responses Using Multiconditional Genetic Interaction Maps. *Mol. Cell* **49**, 346–358 (2013).
  147. Steckel, M. *et al.* Determination of synthetic lethal interactions in KRAS oncogene-dependent cancer cells reveals novel therapeutic targeting strategies. *Cell Res.* **22**, 1227–45 (2012).
  148. Costa-Cabral, S. *et al.* CDK1 Is a synthetic lethal target for KRAS mutant tumours. *PLoS One* **11**, 1–22 (2016).
  149. Unniyampurath, U., Pilankatta, R. & Krishnan, M. N. RNA Interference in the Age of CRISPR: Will CRISPR interfere with RNAi? *Int. J. Mol. Sci.* **17**, 291 (2016).
  150. Lindsay, M. A. Target discovery. *Nat. Rev. Drug Discov.* **2**, 831–838 (2003).
  151. Yang, Y., Adelstein, S. J. & Kassis, A. I. Target discovery from data mining approaches. *Drug Discov. Today* **14**, 147–154 (2009).
  152. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nat. Rev. Drug Discov.* **1**, 727–730 (2002).
  153. Chen, Y. A., Tripathi, L. P. & Mizuguchi, K. TargetMine, an integrated data warehouse for candidate gene prioritisation and target discovery. *PLoS One* **6**, e17844 (2011).
  154. Schmidt, T., Bergner, A. & Schwede, T. Modelling three-dimensional protein structures for applications in drug design. *Drug Discov. Today* **19**, 890–897 (2014).
  155. Song, C. M., Lim, S. J. & Chuantong, J. Recent advances in computer-aided drug design. *Briefings in Bioinformatics* **10**, 579–591 (2009).
  156. Galperin, M. Y., Fernández-Suárez, X. M. & Rigden, D. J. The 24th annual Nucleic Acids Research database issue: a look back and upcoming changes. *Nucleic Acids Res.* **45**, 1–11 (2017).
  157. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, 158–169 (2017).
  158. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
  159. Gaulton, A. *et al.* The ChEMBL database in 2017. *Nucleic Acids Res.* **45**, 945–954 (2017).

160. Tym, J. E. *et al.* canSAR: an updated cancer research and drug discovery knowledgebase. *Nucleic Acids Res.* **44**, 938–943 (2016).
161. Franceschini, A. *et al.* STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41**, 808–815 (2013).
162. Fauman, E. B., Rai, B. K. & Huang, E. S. Structure-based druggability assessment - identifying suitable targets for small molecule therapeutics. *Curr. Opin. Chem. Biol.* **15**, 463–468 (2011).
163. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **23**, 3–25 (1997).
164. Wishart, D. S. *et al.* HMDB 3.0 - The Human Metabolome Database in 2013. *Nucleic Acids Res.* **41**, 801–807 (2013).
165. Le Guilloux, V., Schmidtke, P. & Tuffery, P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* **10**, 168 (2009).
166. Barber, C. B., Dobkin, D. P. & Huhdanpaa, H. The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw.* **22**, 469–483 (1996).
167. Huang, B. & Schroeder, M. LIGSITEcsc: predicting ligand binding sites using the Connolly surface and degree of conservation. *BMC Struct. Biol.* **6**, 19 (2006).
168. Hajduk, P. J., Huth, J. R. & Fesik, S. W. Druggability Indices for Protein Targets Derived from NMR-Based Screening Data. *J. Med. Chem.* **48**, 2518–2525 (2005).
169. Cheng, A. C. *et al.* Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.* **25**, 71–75 (2007).
170. Schmidtke, P. & Barril, X. Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. *J. Med. Chem.* **53**, 5858–5867 (2010).
171. Huang, N. & Jacobson, M. P. Binding-site assessment by virtual fragment screening. *PLoS One* **5**, e10109 (2010).
172. Chen, Y. C. Beware of docking! *Trends Pharmacol. Sci.* **36**, 78–95 (2015).
173. Broomhead, N. K. & Soliman, M. E. Can We Rely on Computational Predictions To Correctly Identify Ligand Binding Sites on Novel Protein Drug Targets? Assessment of Binding Site Prediction Methods and a Protocol for Validation of Predicted Binding Sites. *Cell Biochem. Biophys.* **75**, 15–23 (2017).
174. Seco, J., Luque, F. J. & Barril, X. Binding site detection and druggability index from first principles. *J. Med. Chem.* **52**, 2363–2371 (2009).
175. Schwede, T. Protein modelling: What happened to the ‘protein structure gap’? *Structure* **21**, 1531–1540 (2013).
176. Mullins, J. G. L. Structural modelling pipelines in next generation sequencing projects. *Adv. Protein Chem. Struct. Biol.* **89**, 117–167 (2012).
177. Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
178. Cavasotto, C. N. & Phatak, S. S. Homology modeling in drug discovery: current trends and applications. *Drug Discov. Today* **14**, 676–683 (2009).
179. Biasini, M. *et al.* SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–W258 (2014).

180. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
181. Goujon, M. *et al.* A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* **38**, W695–W699 (2010).
182. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
183. Boratyn, G. M. *et al.* Domain enhanced lookup time accelerated BLAST. *Biol. Direct* **7**, 12 (2012).
184. Blundell, T. L., Sibanda, B. L., Sternberg, M. J. & Thornton, J. M. Knowledge-based prediction of protein structures and the design of novel molecules. *Nature* **326**, 347–352 (1987).
185. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* **31**, 3381–3385 (2003).
186. Accelrys Software Inc. Discovery Studio Modeling Environment. (2013). at <www.accelrys.com>
187. Chemical Computing Group Inc. Molecular Operating Environment (MOE), 2013.08. (2017).
188. Melo, F. & Šali, A. Fold assessment for comparative protein structure modeling. *Protein Sci.* **16**, 2412–2426 (2007).
189. Shen, M.-Y. & Šali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15**, 2507–2524 (2006).
190. Vakser, I. A. Low-resolution structural modeling of protein interactome. *Current Opinion in Structural Biology* **23**, 198–205 (2013).
191. Fan, H. *et al.* Molecular docking screens using comparative models of proteins. *J. Chem. Inf. Model.* **49**, 2512–2527 (2009).
192. Texidó, G. *Genetically Engineered Animal Models for In Vivo Target Identification and Validation in Oncology. Target Identification and Validation in Drug Discovery: Methods and Protocols, Methods in Molecular Biology* **986**, (Springer Science and Business Media New York, 2013).
193. Costantini, F. & Lacy, E. Introduction of a rabbit  $\beta$ -globin gene into the mouse germ line. *Nature* **294**, 92–94 (1981).
194. Capecchi, M. R. Altering the Genome Homologous Recombination. *Science (80-. )*. **244**, 1288–1292 (1989).
195. Abell, A. N. *et al.* Ablation of MEKK4 kinase activity causes neurulation and skeletal patterning defects in the mouse embryo. *Mol. Cell. Biol.* **25**, 8948–8959 (2005).
196. Puyol, M. *et al.* A Synthetic Lethal Interaction between K-Ras Oncogenes and Cdk4 Unveils a Therapeutic Strategy for Non-small Cell Lung Carcinoma. *Cancer Cell* **18**, 63–73 (2010).
197. Degrassi, A. *et al.* Efficacy of PHA-848125, a cyclin-dependent kinase inhibitor, on the K-Ras(G12D)LA2 lung adenocarcinoma transgenic mouse model: evaluation by multimodality imaging. *Mol. Cancer Ther.* **9**, 673–681 (2010).
198. Austin, C. P. *et al.* The Knockout Mouse Project. *Nat. Genet.* **36**, 921–924 (2004).
199. Fell, A. NIH renews UC Davis Knockout Mouse Project. *Scienmag* (2016). at

- <<https://scienmag.com/nih-renews-uc-davis-knockout-mouse-project/>>
200. Sauer, B. & Henderson, N. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 5166–5170 (1988).
  201. Gaj, T., Sirk, S. J., Shui, S. & Liu, J. Genome-Editing Technologies: Principles and Applications. *Cold Spring Harb. Perspect. Biol.* **8**, (2016).
  202. Editorial. Method of the Year 2011. *Nat. Methods* **9**, 1 (2012).
  203. Mazzeletti, M. & Texidó, G. *In Vivo Target Validation by Inducible RNAi in Human Xenograft Mouse Models. Target Identification and Validation in Drug Discovery: Methods and Protocols, Methods in Molecular Biology* **986**, (Springer Science and Business Media New York, 2013).
  204. Barrès, V. *et al.* An essential role for Ran GTPase in epithelial ovarian cancer cell survival. *Mol. Cancer* **9**, 272 (2010).
  205. Lee-Hoeflich, S. T. *et al.* A central role for HER3 in HER2-amplified breast cancer: Implications for targeted therapy. *Cancer Res.* **68**, 5878–5887 (2008).
  206. Sala, G. *et al.* Phospholipase C $\gamma$ 1 is required for metastasis development and progression. *Cancer Res.* **68**, 10187–10196 (2008).
  207. Kleinhammer, A., Wurst, W. & Kühn, R. *Target Validation in Mice by Constitutive and Conditional RNAi. Target Identification and Validation in Drug Discovery: Methods and Protocols, Methods in Molecular Biology* **986**, (Springer Science and Business Media New York, 2013).
  208. Bobbin, M. L. & Rossi, J. J. RNA Interference (RNAi)-Based Therapeutics: Delivering on the Promise? *Annu. Rev. Pharmacol. Toxicol.* **56**, 103–122 (2016).
  209. Barata, P., Sood, A. K. & Hong, D. S. RNA-targeted therapeutics in cancer clinical trials: Current status and future directions. *Cancer Treat. Rev.* **50**, 35–47 (2016).
  210. Honore, P. *et al.* Analgesic profile of intrathecal P2X $_3$  antisense oligonucleotide treatment in chronic inflammatory and neuropathic pain states in rats. *Pain* **99**, 11–19 (2002).
  211. Zhang, N., Chang, Y., Rios, A. & An, Z. HER3/ErbB3, an emerging cancer therapeutic target. *Acta Biochim. Biophys. Sin. (Shanghai)*. **48**, 39–48 (2016).
  212. Zanders, E. D., Bailey, D. S. & Dean, P. M. Probes for chemical genomics by design. *Drug Discov. Today* **7**, 711–718 (2002).
  213. Editorial. Stay on target. *Nat. Chem. Biol.* **9**, 193 (2013).
  214. Bunnage, M. E., Chekler, E. L. P. & Jones, L. H. Target validation using chemical probes. *Nat. Publ. Gr.* **9**, 195–199 (2013).
  215. Workman, P. & Collins, I. Probing the Probes: Fitness Factors For Small Molecule Tools. *Chem. Biol.* **17**, 561–577 (2010).
  216. Blagg, J. & Workman, P. Chemical biology approaches to target validation in cancer. *Curr. Opin. Pharmacol.* **17**, 87–100 (2014).
  217. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010).
  218. Nicodeme, E. *et al.* Suppression of inflammation by a synthetic histone mimic. *Nature* **468**, 1119–1123 (2010).



219. Liu, L., Zhen, X. T., Denton, E., Marsden, B. D. & Schapira, M. ChromoHub: A data hub for navigators of chromatin-mediated signalling. *Bioinformatics* **28**, 2205–2206 (2012).
220. Valdar, W. S. J. Scoring Residue Conservation. *Proteins* **48**, 227–241 (2002).
221. Kelly, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modelling, prediction, and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
222. Haering, C. H., Löwe, J., Hochwagen, A. & Nasmyth, K. Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol. Cell* **9**, 773–788 (2002).
223. Alt, A. *et al.* Specialized interfaces of Smc5/6 control hinge stability and DNA association. *Nat. Commun.* **8**, 14011 (2017).
224. Lammens, A., Schele, A. & Hopfner, K.-P. Structural Biochemistry of ATP-Driven Dimerization and DNA-Stimulated Activation of SMC ATPases. *Curr. Biol.* **14**, 1778–1782 (2004).
225. Doyle, J. M., Gao, J., Wang, J., Yang, M. & Potts, P. R. MAGE-RING Protein Complexes Comprise a Family of E3 Ubiquitin Ligases. *Mol. Cell* **39**, 963–974 (2010).
226. Duan, X. *et al.* Structural and Functional Insights into the Roles of the Mms21 Subunit of the Smc5/6 Complex. *Mol. Cell* **35**, 657–668 (2009).
227. Newman, J. A. *et al.* Structures of two melanoma-associated antigens suggest allosteric regulation of effector binding. *PLoS One* **11**, e0148762 (2016).
228. Kurze, A. *et al.* A positively charged channel within the Smc1/Smc3 hinge required for sister chromatid cohesion. *EMBO J.* **30**, 364–378 (2011).
229. Pellegrino, S. *et al.* Structural and functional characterization of an SMC-like protein RecN: New insights into double-strand break repair. *Structure* **20**, 2076–2086 (2012).
230. Brooks, B. R. *et al.* CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **4**, 187–217 (1983).
231. Brooks, B. R. *et al.* CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **30**, 1545–1614 (2009).
232. The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.
233. Haering, C. H. *et al.* Structure and stability of cohesin's Smc1-kleisin interaction. *Mol. Cell* **15**, 951–964 (2004).
234. Liew, C. W., Sun, H., Hunter, T. & Day, C. L. RING domain dimerization is essential for RNF4 function. *Biochem. J.* **431**, 23–29 (2010).
235. Plechanovová, A. *et al.* Mechanism of ubiquitylation by dimeric RING ligase RNF4. *Nat. Struct. Mol. Biol.* **18**, 1052–1059 (2011).
236. Natsume, R. *et al.* Structure and function of the histone chaperone CIA/ASF1 complexed with histones H3 and H4. *Nature* **446**, 338–341 (2007).
237. Tang, Y. *et al.* Structure of a human ASF1a-HIRA complex and insights into specificity of histone chaperone complex assembly. *Nat. Struct. Mol. Biol.* **13**, 921–929 (2006).
238. Filippakopoulos, P. *et al.* Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. *Cell* **149**, 214–231 (2012).
239. Owen, D. J. *et al.* The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase Gcn5p. *EMBO J.* **19**, 6141–6149 (2000).
240. Chaikuad, A. *et al.* Structure-Based Identification of Inhibitory Fragments Targeting the

- p300/CBP-Associated Factor Bromodomain. *J. Med. Chem.* **59**, 1648–1653 (2016).
241. Schuetz, A. *et al.* Crystal Structure of a Binary Complex Between Human GCN5 Histone Acetyltransferase Domain and Acetyl Coenzyme A. *Proteins* **68**, 403–407 (2007).
  242. Berthold, M. R. *et al.* *KNIME: The Konstanz Information Miner. Data Analysis, Machine Learning and Applications* (Springer Berlin Heidelberg, 2008). doi:10.1007/978-3-540-78246-9\_38
  243. GraphPad Software Inc. GraphPad Prism, version 7.02 for Windows, La Jolla California USA. (2017). at <www.graphpad.com>
  244. Dotmatics Limited. Dotmatics Suite, version 5.1, Bishops Stortford UK. (2017). at <www.dotmatics.com>
  245. Uhlmann, F. SMC complexes: from DNA to chromosomes. *Nat. Rev. Mol. Cell Biol.* **17**, 399–412 (2016).
  246. Niki, H., Jaffé, A., Imamura, R., Ogura, T. & Hiraga, S. The new gene mukB codes for a 177 kd protein with coiled-coil domains involved in chromosome partitioning of E.coli. *EMBO J.* **10**, 183–193 (1991).
  247. Strunnikov, A. V., Larionov, V. L. & Koshland, D. SMC1: An Essential Yeast Gene Encoding a Putative Head-Rod-Tail Protein Is Required for Nuclear Division and Defines a New Ubiquitous Protein Family. *J. Cell Biol.* **123**, 1635–1648 (1993).
  248. Hirano, T. & Mitchison, T. J. A heterodimeric coiled-coil protein required for mitotic chromosome condensation in vitro. *Cell* **79**, 449–458 (1994).
  249. Lehmann, A. R. *et al.* The rad18 gene of *Schizosaccharomyces pombe* defines a new subgroup of the SMC superfamily involved in DNA repair. *Mol. Cell. Biol.* **15**, 7067–7080 (1995).
  250. Losada, A. & Hirano, T. Dynamic molecular linkers of the genome: The first decade of SMC proteins. *Genes Dev.* **19**, 1269–1287 (2005).
  251. Wu, N. & Yu, H. The Smc complexes in DNA damage response. *Cell Biosci.* **2**, 5 (2012).
  252. Nasmyth, K. Segregating Sister Genomes: The Molecular Biology of Chromosome Separation. *Science (80-. )*. **297**, 559–565 (2002).
  253. Hirano, T. Condensins: Organizing and segregating the genome. *Curr. Biol.* **15**, 265–275 (2005).
  254. Ampatzidou, E., Irmisch, A., O’Connell, M. J. & Murray, J. M. Smc5/6 is required for repair at collapsed replication forks. *Mol. Cell. Biol.* **26**, 9387–9401 (2006).
  255. De Piccoli, G. *et al.* Smc5-Smc6 mediate DNA double-strand-break repair by promoting sister-chromatid recombination. *Nat. Cell Biol.* **8**, 1032–1034 (2006).
  256. Murray, J. M. & Carr, A. M. Smc5/6: a link between DNA repair and unidirectional replication? *Nat. Rev. Mol. Cell Biol.* **9**, 177–182 (2008).
  257. Verkade, H. M., Bugg, S. J., Lindsay, H. D., Carr, A. M. & O’Connell, M. J. Rad18 is required for DNA repair and checkpoint responses in fission yeast. *Mol. Biol. Cell* **10**, 2905–2918 (1999).
  258. Tedeschi, A. *et al.* Wapl is an essential regulator of chromatin structure and chromosome segregation. *Nature* **501**, 564–568 (2013).
  259. Aono, N., Sutani, T., Tomonaga, T., Mochida, S. & Yanagida, M. Cnd2 has dual roles in mitotic condensation and interphase. *Nature* **417**, 197–202 (2002).

260. Birkenbihl, R. P. & Subramani, S. Cloning and characterization of rad21 an essential gene of *Schizosaccharomyces pombe* involved in DNA double-strand-break repair. *Nucleic Acids Res.* **20**, 6605–6611 (1992).
261. Melby, T. E., Ciampaglio, C. N., Briscoe, G. & Erickson, H. P. The symmetrical structure of structural maintenance of chromosomes (SMC) and MukB proteins: Long, antiparallel coiled coils, folded at a flexible hinge. *J. Cell Biol.* **142**, 1595–1604 (1998).
262. Hirano, M. & Hirano, T. Positive and negative regulation of SMC-DNA interactions by ATP and accessory proteins. *EMBO J.* **23**, 2664–2673 (2004).
263. Bürmann, F. *et al.* An asymmetric SMC-kleisin bridge in prokaryotic condensin. *Nat. Struct. Mol. Biol.* **20**, 371–379 (2013).
264. Gligoris, T. G. *et al.* Closing the cohesin ring: structure and function of its Smc3-kleisin interface. *Science (80-. )*. **346**, 963–967 (2014).
265. Anderson, D. E., Losada, A., Erickson, H. P. & Hirano, T. Condensin and cohesin display different arm conformations with characteristic hinge angles. *J. Cell Biol.* **156**, 419–424 (2002).
266. Haering, C. H., Farcas, A.-M., Arumugam, P., Metson, J. & Nasmyth, K. The cohesin ring concatenates sister DNA molecules. *Nature* **454**, 297–301 (2008).
267. Cuylen, S., Metz, J. & Haering, C. H. Condensin structures chromosomal DNA through topological links. *Nat. Struct. Mol. Biol.* **18**, 894–901 (2011).
268. Kanno, T., Berta, D. G. & Sjögren, C. The Smc5/6 Complex Is an ATP-Dependent Intermolecular DNA Linker. *Cell Rep.* **12**, 1471–1482 (2015).
269. Gerlich, D., Koch, B., Dupeux, F., Peters, J.-M. & Ellenberg, J. Live-cell imaging reveals a stable cohesin-chromatin interaction after but not before DNA replication. *Curr. Biol.* **16**, 1571–1578 (2006).
270. Gruber, S. *et al.* Evidence that Loading of Cohesin Onto Chromosomes Involves Opening of Its SMC Hinge. *Cell* **127**, 523–537 (2006).
271. Buheitel, J. & Stemmann, O. Prophase pathway-dependent removal of cohesin from human chromosomes requires opening of the Smc3-Scc1 gate. *EMBO J.* **32**, 666–676 (2013).
272. Murayama, Y. & Uhlmann, F. Biochemical reconstitution of topological DNA binding by the cohesin ring. *Nature* **505**, 367–371 (2014).
273. Murayama, Y. & Uhlmann, F. DNA Entry into and Exit out of the Cohesin Ring by an Interlocking Gate Mechanism. *Cell* **163**, 1628–1640 (2015).
274. Huis in 't Veld, P. J. *et al.* Characterization of a DNA exit gate in the human cohesin ring. *Science (80-. )*. **346**, 968–972 (2014).
275. Yoshimura, S. H. *et al.* Condensin Architecture and Interaction with DNA: Regulatory Non-SMC Subunits Bind to the Head of SMC Heterodimer. *Curr. Biol.* **12**, 508–513 (2002).
276. Sakai, A., Hizume, K., Sutani, T., Takeyasu, K. & Yanagida, M. Condensin but not cohesin SMC heterodimer induces DNA reannealing through protein-protein assembly. *EMBO J.* **22**, 2764–2775 (2003).
277. Palecek, J., Vidot, S., Feng, M., Doherty, A. J. & Lehmann, A. R. The Smc5-Smc6 DNA repair complex: Bridging of the Smc5-Smc6 heads by the kleisin, Nse4, and non-kleisin subunits. *J. Biol. Chem.* **281**, 36952–36959 (2006).

278. Duan, X. *et al.* Architecture of the Smc5/6 complex of *Saccharomyces cerevisiae* reveals a unique interaction between the Nse5-6 subcomplex and the hinge regions of Smc5 and Smc6. *J. Biol. Chem.* **284**, 8507–8515 (2009).
279. Huber, R. G. *et al.* Impairing Cohesin Smc1/3 Head Engagement Compensates for the Lack of Eco1 Function. *Structure* **24**, 1991–1999 (2016).
280. Gruber, S. Shaping chromosomes by DNA capture and release: gating the SMC rings. *Curr. Opin. Cell Biol.* **46**, 87–93 (2017).
281. Ocampo-Hafalla, M. T., Katou, Y., Shirahige, K. & Uhlmann, F. Displacement and re-accumulation of centromeric cohesin during transient pre-anaphase centromere splitting. *Chromosoma* **116**, 531–544 (2007).
282. D'Ambrosio, C. *et al.* Identification of cis-acting sites for condensin loading onto budding yeast chromosomes. *Genes Dev.* **22**, 2215–2227 (2008).
283. Jeppsson, K. *et al.* The Chromosomal Association of the Smc5/6 Complex Depends on Cohesion and Predicts the Level of Sister Chromatid Entanglement. *PLoS Genet.* **10**, e1004680 (2014).
284. Natsume, T. *et al.* Kinetochores coordinate pericentromeric cohesion and early DNA replication by Cdc7-Dbf4 kinase recruitment. *Mol. Cell* **50**, 661–674 (2013).
285. Verzijlbergen, K. F. *et al.* Shugoshin biases chromosomes for biorientation through condensin recruitment to the pericentromere. *Elife* **3**, e01374 (2014).
286. Kagey, M. H. *et al.* Mediator and Cohesin Connect Gene Expression and Chromatin Architecture. *Nature* **467**, 430–435 (2010).
287. Kranz, A.-L. *et al.* Genome-wide analysis of condensin binding in *Caenorhabditis elegans*. *Genome Biol.* **14**, R112 (2013).
288. Lopez-Serra, L., Kelly, G., Patel, H., Stewart, A. & Uhlmann, F. The Scc2–Scc4 complex acts in sister chromatid cohesion and transcriptional regulation by maintaining nucleosome-free regions. *Nat. Genet.* **46**, 1147–1151 (2014).
289. Haeusler, R. A., Pratt-Hyatt, M., Good, P. D., Gipson, T. A. & Engelke, D. R. Clustering of yeast tRNA genes is mediated by specific association of condensin with tRNA gene transcription complexes. *Genes Dev.* **22**, 2204–2214 (2008).
290. Johzuka, K. & Horiuchi, T. The cis Element and Factors Required for Condensin Recruitment to Chromosomes. *Mol. Cell* **34**, 26–35 (2009).
291. Lengronne, A. *et al.* Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* **430**, 573–578 (2004).
292. Schmidt, C. K., Brookes, N. & Uhlmann, F. Conserved features of cohesin binding along fission yeast chromosomes. *Genome Biol.* **10**, R52 (2009).
293. Parelho, V. *et al.* Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. *Cell* **132**, 422–433 (2008).
294. Ong, C.-T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.* **15**, 234–246 (2014).
295. Deardorff, M. A. *et al.* HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. *Nature* **489**, 313–317 (2012).
296. Sutani, T. *et al.* Condensin targets and reduces unwound DNA structures associated with transcription in mitotic chromosome condensation. *Nat. Commun.* **6**, 7815 (2015).

297. Remeseiro, S., Cuadrado, A., Gómez-López, G., Pisano, D. G. & Losada, A. A unique role of cohesin-SA1 in gene regulation and development. *EMBO J.* **31**, 2090–2102 (2012).
298. Schleiffer, A. *et al.* Kleisins: A superfamily of bacterial and eukaryotic SMC protein partners. *Mol. Cell* **11**, 571–575 (2003).
299. Hadjur, S. *et al.* Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* **460**, 410–413 (2009).
300. Chao, W. C. H. *et al.* Structural Studies Reveal the Functional Modularity of the Scc2-Scc4 Cohesin Loader. *Cell Rep.* **12**, 719–725 (2015).
301. Zhang, J. *et al.* Acetylation of Smc3 by Eco1 Is Required for S Phase Sister Chromatid Cohesion in Both Human and Yeast. *Mol. Cell* **31**, 143–151 (2008).
302. Nishiyama, T. *et al.* Sororin mediates sister chromatid cohesion by antagonizing Wapl. *Cell* **143**, 737–749 (2010).
303. Çamdere, G., Guacci, V., Stricklin, J. & Koshland, D. The ATPases of cohesin interface with regulators to modulate cohesin-mediated DNA tethering. *Elife* **4**, e11315 (2015).
304. Song, J. *et al.* Cohesin acetylation promotes sister chromatid cohesion only in association with the replication machinery. *J. Biol. Chem.* **287**, 34325–34336 (2012).
305. Minamino, M. *et al.* Esco1 Acetylates Cohesin via a Mechanism Different from That of Esco2. *Curr. Biol.* **25**, 1694–1706 (2015).
306. Nasmyth, K. & Haering, C. H. Cohesin: Its Roles and Mechanisms. *Annu. Rev. Genet.* **43**, 525–558 (2009).
307. Gandhi, R., Gillespie, P. P. J. & Hirano, T. Human Wapl Is a Cohesin-Binding Protein that Promotes Sister-Chromatid Resolution in Mitotic Prophase. *Curr. Biol.* **16**, 2406–2417 (2006).
308. Nishiyama, T., Sykora, M. M., Huis In 't Veld, P. J., Mechtler, K. & Peters, J.-M. Aurora B and Cdk1 mediate Wapl activation and release of acetylated cohesin from chromosomes by phosphorylating Sororin. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 13404–13409 (2013).
309. Uhlmann, F., Wernic, D., Poupart, M.-A., Koonin, E. V. & Nasmyth, K. Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast. *Cell* **103**, 375–386 (2000).
310. Bauerschmidt, C. *et al.* Cohesin promotes the repair of ionizing radiation-induced DNA double-strand breaks in replicated chromatin. *Nucleic Acids Res.* **38**, 477–487 (2010).
311. Kim, J., Krasieva, T. B., LaMorte, V., Taylor, A. M. R. & Yokomori, K. Specific recruitment of human cohesin to laser-induced DNA damage. *J. Biol. Chem.* **277**, 45149–45153 (2002).
312. Nagao, K., Adachi, Y. & Yanagida, M. Separase-mediated cleavage of cohesin at interphase is required for DNA repair. *Nature* **430**, 1044–1048 (2004).
313. Yazdi, P. T. *et al.* SMC1 is a downstream effector in the ATM/NBS1 branch of the human S-phase checkpoint. *Genes Dev.* **16**, 571–582 (2002).
314. Watrin, E. & Peters, J. The cohesin complex is required for the DNA damage-induced G2/M checkpoint in mammalian cells. *EMBO J.* **28**, 2625–2635 (2009).
315. Neuwald, A. F. & Hirano, T. HEAT Repeats Associated with Condensins , Cohesins , and Other Complexes Involved in Chromosome-Related Functions. *Genome Res.* **10**, 1445–1452 (2000).

316. Piazza, I. *et al.* Association of condensin with chromosomes depends on DNA binding by its HEAT-repeat subunits. *Nat. Struct. Mol. Biol.* **21**, 560–568 (2014).
317. Onn, I., Aono, N., Hirano, M. & Hirano, T. Reconstitution and subunit geometry of human condensin complexes. *EMBO J.* **26**, 1024–1034 (2007).
318. Barysz, H. *et al.* Three-dimensional topology of the SMC2/SMC4 subcomplex from chicken condensin I revealed by cross-linking and molecular modelling. *Open Biol.* **5**, 150005 (2015).
319. Soh, Y.-M. *et al.* Molecular basis for SMC rod formation and its dissolution upon DNA binding. *Mol. Cell* **57**, 290–303 (2015).
320. Hirota, T., Gerlich, D., Koch, B., Ellenberg, J. & Peters, J.-M. Distinct functions of condensin I and II in mitotic chromosome assembly. *J. Cell Sci.* **117**, 6435–6445 (2004).
321. Kalitsis, P., Zhang, T., Marshall, K. M., Nielsen, C. F. & Hudson, D. F. Condensin, master organizer of the genome. *Chromosom. Res.* **25**, 61–76 (2017).
322. Zhang, T. *et al.* Condensin I and II behaviour in interphase nuclei and cells undergoing premature chromosome condensation. *Chromosom. Res.* **24**, 243–269 (2016).
323. Hirano, T. Condensins: Universal organizers of chromosomes with diverse functions. *Genes Dev.* **26**, 1659–1678 (2012).
324. Green, L. C. *et al.* Contrasting roles of condensin I and condensin II in mitotic chromosome formation. *J. Cell Sci.* **125**, 1591–1604 (2012).
325. Hudson, D. F. *et al.* Molecular and Genetic Analysis of Condensin Function in Vertebrate Cells. *Mol. Biol. Cell* **19**, 3070–3079 (2008).
326. Kinoshita, K., Kobayashi, T. J. & Hirano, T. Balancing acts of two HEAT subunits of condensin I support dynamic assembly of chromosome axes. *Dev. Cell* **33**, 94–107 (2015).
327. Akai, Y. *et al.* Opposing role of condensin hinge against replication protein A in mitosis and interphase through promoting DNA annealing. *Open Biol.* **1**, 110023 (2011).
328. Kimura, K., Rybenkov, V. V., Crisona, N. J., Hirano, T. & Cozzarelli, N. R. 13S condensin actively reconfigures DNA by introducing global positive writhe: Implications for chromosome condensation. *Cell* **98**, 239–248 (1999).
329. Gilbert, N. & Allan, J. Supercoiling in DNA and chromatin. *Curr. Opin. Genet. Dev.* **25**, 15–21 (2014).
330. Farr, C. J., Antoniou-Kourounioti, M., Mimmack, M. L., Volkov, A. & Porter, A. C. G. The  $\alpha$  isoform of topoisomerase II is required for hypercompaction of mitotic chromosomes in human cells. *Nucleic Acids Res.* **42**, 4414–4426 (2014).
331. Charbin, A., Bouchoux, C. & Uhlmann, F. Condensin aids sister chromatid decatenation by topoisomerase II. *Nucleic Acids Res.* **42**, 340–348 (2014).
332. Kimura, K., Cuvier, O. & Hirano, T. Chromosome Condensation by a Human Condensin Complex in *Xenopus* Egg Extracts. *J. Biol. Chem.* **276**, 5417–5420 (2001).
333. Baxter, J. *et al.* Positive supercoiling of mitotic DNA drives decatenation by topoisomerase II in eukaryotes. *Science (80-. ).* **331**, 1328–1332 (2011).
334. Hirano, T. At the heart of the chromosome: SMC proteins in action. *Nat. Rev. Mol. Cell Biol.* **7**, 311–322 (2006).
335. Poonperm, R. *et al.* Chromosome Scaffold is a Double-Stranded Assembly of Scaffold

- Proteins. *Sci. Rep.* **5**, 11916 (2015).
336. König, P., Braunfeld, M. B., Sedat, J. W. & Agard, D. A. The three-dimensional structure of in vitro reconstituted *Xenopus laevis* chromosomes by EM tomography. *Chromosoma* **116**, 349–372 (2007).
  337. Cheng, T. M. K. *et al.* A simple biophysical model emulates budding yeast chromosome condensation. *Elife* **4**, e05565 (2015).
  338. Goloborodko, A., Imakaev, M. V., Marko, J. F. & Mirny, L. Compaction and segregation of sister chromatids via active loop extrusion. *Elife* **5**, 14864 (2016).
  339. Heale, J. T. *et al.* Condensin I interacts with the PARP-1-XRCC1 complex and functions in DNA single-strand break repair. *Mol. Cell* **21**, 837–848 (2006).
  340. Wood, J. L., Liang, Y., Li, K. & Chen, J. Microcephalin/MCPH1 associates with the condensin II complex to function in homologous recombination repair. *J. Biol. Chem.* **283**, 29586–29592 (2008).
  341. Kwan, C., Wei, Y. & Zheng, X. F. S. Compacting DNA During the Interphase: Condensin Maintains rDNA Integrity. *Cell Cycle* **6**, 2213–2218 (2007).
  342. De Piccoli, G., Torres-Rosell, J. & Aragón, L. The unnamed complex: What do we know about Smc5-Smc6? *Chromosom. Res.* **17**, 251–263 (2009).
  343. Taylor, E. M., Copsey, A. C., Hudson, J. J. R., Vidot, S. & Lehmann, A. R. Identification of the proteins, including MAGEG1, that make up the human SMC5-6 protein complex. *Mol. Cell. Biol.* **28**, 1197–1206 (2008).
  344. Fujioka, Y., Kimata, Y., Nomaguchi, K., Watanabe, K. & Kohno, K. Identification of a novel non-structural maintenance of chromosomes (SMC) component of the SMC5-SMC6 complex involved in DNA repair. *J. Biol. Chem.* **277**, 21585–21591 (2002).
  345. Palecek, J. J. & Gruber, S. Kite Proteins: A Superfamily of SMC/Kleisin Partners Conserved Across Bacteria, Archaea, and Eukaryotes. *Structure* **23**, 2183–2190 (2015).
  346. Sergeant, J. *et al.* Composition and architecture of the *Schizosaccharomyces pombe* Rad18 (Smc5-6) complex. *Mol. Cell. Biol.* **25**, 172–184 (2005).
  347. Andrews, E. A. *et al.* Nse2, a component of the Smc5-6 complex, is a SUMO ligase required for the response to DNA damage. *Mol. Cell. Biol.* **25**, 185–196 (2005).
  348. Nasim, A. & Smith, B. P. Genetic control of radiation sensitivity in *Schizosaccharomyces Pombe*. *Genetics* **79**, 573–582 (1975).
  349. Pebernard, S., Perry, J. J. P., Tainer, J. A. & Boddy, M. N. Nse1 RING-like domain supports functions of the Smc5-Smc6 holocomplex in genome stability. *Mol. Biol. Cell* **19**, 4099–4109 (2008).
  350. Bermúdez-López, M. *et al.* The Smc5/6 complex is required for dissolution of DNA-mediated sister chromatid linkages. *Nucleic Acids Res.* **38**, 6502–6512 (2010).
  351. Torres-Rosell, J. *et al.* The Smc5-Smc6 complex and SUMO modification of Rad52 regulates recombinational repair at the ribosomal gene locus. *Nat. Cell Biol.* **9**, 923–931 (2007).
  352. Kegel, A. & Sjögren, C. The Smc5/6 complex: More than repair? *Cold Spring Harb. Symp. Quant. Biol.* **75**, 179–187 (2010).
  353. Gallego-Paez, L. M. *et al.* Smc5/6-mediated regulation of replication progression contributes to chromosome assembly during mitosis in human cells. *Mol. Biol. Cell* **25**, 302–317 (2014).

354. Menolfi, D., Delamarre, A., Lengronne, A., Pasero, P. & Branzei, D. Essential Roles of the Smc5/6 Complex in Replication through Natural Pausing Sites and Endogenous DNA Damage Tolerance. *Mol. Cell* **60**, 835–846 (2015).
355. Farcas, A.-M., Uluocak, P., Helmhart, W. & Nasmyth, K. Cohesin's Concatenation of Sister DNAs Maintains Their Intertwining. *Mol. Cell* **44**, 97–107 (2011).
356. Peng, J. & Feng, W. Incision of damaged DNA in the presence of an impaired Smc5/6 complex imperils genome stability. *Nucleic Acids Res.* **44**, 10216–10229 (2016).
357. Roy, M. A., Siddiqui, N. & D'Amours, D. Dynamic and selective DNA-binding activity of Smc5, a core component of the Smc5-Smc6 complex. *Cell Cycle* **10**, 690–700 (2011).
358. Roy, M. A. & D'Amours, D. DNA-binding properties of Smc6, a core component of the Smc5-6 DNA repair complex. *Biochem. Biophys. Res. Commun.* **416**, 80–85 (2011).
359. Fousteri, M. I. & Lehmann, A. R. A novel SMC protein complex in *Schizosaccharomyces pombe* contains the Rad18 DNA repair protein. *EMBO J.* **19**, 1691–1702 (2000).
360. Zabradý, K. *et al.* Chromatin association of the SMC5/6 complex is dependent on binding of its NSE3 subunit to DNA. *Nucleic Acids Res.* **44**, 1064–1079 (2015).
361. Potts, P. R. & Yu, H. The SMC5/6 complex maintains telomere length in ALT cancer cells through SUMOylation of telomere-binding proteins. *Nat. Struct. Mol. Biol.* **14**, 581–590 (2007).
362. Blackburn, E. H. Switching and Signaling at the Telomere. *Cell* **106**, 661–673 (2001).
363. Harley, C. B., Futcher, A. B. & Greider, C. W. Telomeres shorten during ageing of human fibroblasts. *Nature* **345**, 458–460 (1990).
364. Bodnar, A. G. *et al.* Extension of life-span by introduction of telomerase into normal human cells. *Science (80-. )*. **279**, 349–352 (1998).
365. Shay, J. W. & Bacchetti, S. A survey of telomerase activity in human cancer. *Eur. J. Cancer* **33**, 787–791 (1997).
366. Henson, J. D., Neumann, A. A., Yeager, T. R. & Reddel, R. R. Alternative Lengthening of Telomeres in Mammalian Cells. *Oncogene* **21**, 598–610 (2002).
367. Dunham, M. A., Neumann, A. A., Fasching, C. L. & Reddel, R. R. Telomere maintenance by recombination in human cells. *Nat. Genet.* **26**, 447–450 (2000).
368. Yeager, T. R. *et al.* Telomerase-negative Immortalized Human Cells Contain a Novel Type of Promyelocytic Leukemia (PML) Body. *Cancer Res.* **59**, 4175–4179 (1999).
369. De Lange, T. Shelterin: The protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19**, 2100–2110 (2005).
370. Chavez, A., George, V., Agrawal, V. & Johnson, F. B. Sumoylation and the structural maintenance of chromosomes (Smc) 5/6 complex slow senescence through recombination intermediate resolution. *J. Biol. Chem.* **285**, 11922–11930 (2010).
371. Moradi-Fard, S. *et al.* Smc5/6 Is a Telomere-Associated Complex that Regulates Sir4 Binding and TPE. *PLoS Genet.* **12**, e1006268 (2016).
372. Agostino, M. *Introduction to the BLAST Suite and BLASTN. Practical Bioinformatics* (Garland Science, 2012). at  
<[http://www.garlandscience.com/res/pdf/practicalbioinformatics\\_ch3.pdf](http://www.garlandscience.com/res/pdf/practicalbioinformatics_ch3.pdf)>
373. Scott, D. E. *et al.* Using a Fragment-Based Approach To Target Protein-Protein Interactions. *ChemBioChem* **14**, 332–342 (2013).



374. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127–134 (1995).
375. Yin, H. & Hamilton, A. D. Strategies for targeting protein-protein interactions with synthetic agents. *Angew. Chemie - Int. Ed.* **44**, 4130–4163 (2005).
376. Hodges, C., Kirkland, J. G. & Crabtree, G. R. The many roles of BAF (mSWI/SNF) and PBAF complexes in cancer. *Cold Spring Harb. Perspect. Med.* **6**, (2016).
377. Xia, W. *et al.* BAF180 is a critical regulator of p21 induction and a tumor suppressor mutated in breast cancer. *Cancer Res.* **68**, 1667–1674 (2008).
378. Kschonsak, M. & Haering, C. H. Shaping mitotic chromosomes: From classical concepts to molecular mechanisms. *BioEssays* **37**, 755–766 (2015).
379. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* **319**, 1097–1113 (2002).
380. Arrowsmith, C. H., Bountra, C., Fish, P. V., Lee, K. & Schapira, M. Epigenetic protein families: a new frontier for drug discovery. *Nat. Rev. Drug Discov.* **11**, 384–400 (2012).
381. Roth, S. Y., Denu, J. M. & Allis, C. D. Histone Acetyltransferases. *Annu. Rev. Biochem.* **70**, 81–120 (2001).
382. Wilson, B. G. & Roberts, C. W. M. SWI/SNF nucleosome remodellers and cancer. *Nat. Rev. Cancer* **11**, 481–492 (2011).
383. Shain, A. H. & Pollack, J. R. The Spectrum of SWI/SNF Mutations, Ubiquitous in Human Cancers. *PLoS One* **8**, e55119 (2013).
384. Kadoch, C. *et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* **45**, 592–601 (2013).
385. Stern, M., Jensen, R. & Herskowitz, I. Five SWI genes are required for expression of the HO gene in yeast. *J. Mol. Biol.* **178**, 853–868 (1984).
386. Carlson, M., Osmond, B. C. & Botstein, D. Mutants of yeast defective in sucrose utilization. *Genetics* **98**, 25–40 (1981).
387. Cairns, B. R., Kim, Y.-J., Sayre, M. H., Laurent, B. C. & Kornberg, R. D. A multisubunit complex containing the SWI1/ADR6, SWI2/SNF2, SWI3, SNF5, and SNF6 gene products isolated from yeast. *Proc. Natl. Acad. Sci.* **91**, 1950–1954 (1994).
388. Peterson, C. L., Dingwall, A. & Scottt, M. P. Five SWI/SNF gene products are components of a large multisubunit complex required for transcriptional enhancement. *Proc. Natl. Acad. Sci.* **91**, 2905–2908 (1994).
389. Narlikar, G. J., Sundaramoorthy, R. & Owen-Hughes, T. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* **154**, 490–503 (2013).
390. Saha, A., Wittmeyer, J. & Cairns, B. R. Chromatin remodelling: the industrial revolution of DNA around histones. *Nat. Rev. Mol. Cell Biol.* **7**, 437–447 (2006).
391. Gong, F., Fahy, D. & Smerdon, M. J. Rad4–Rad23 interaction with SWI/SNF links ATP-dependent chromatin remodeling with nucleotide excision repair. *Nat. Struct. Mol. Biol.* **13**, 902–907 (2006).
392. Chai, B., Huang, J., Cairns, B. R. & Laurent, B. C. Distinct roles for the RSC and Swi/Snf ATP-dependent chromatin remodelers in DNA double-strand break repair. *Genes Dev.* **19**, 1656–1661 (2005).

393. Ho, L. & Crabtree, G. R. Chromatin remodelling during development. *Nature* **463**, 474–484 (2010).
394. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
395. Bao, X. *et al.* A novel ATAC-seq approach reveals lineage-specific reinforcement of the open chromatin landscape via cooperation between BAF and p63. *Genome Biol.* **16**, 284 (2015).
396. Lemon, B., Inouye, C., King, D. S. & Tjian, R. Selectivity of chromatin-remodelling cofactors for ligand-activated transcription. *Nature* **414**, 924–928 (2001).
397. Xue, Y. *et al.* The human SWI/SNF-B chromatin-remodeling complex is related to yeast Rsc and localizes at kinetochores of mitotic chromosomes. *Proc. Natl. Acad. Sci.* **97**, 13015–13020 (2000).
398. Bajpai, R. *et al.* CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature* **463**, 958–962 (2010).
399. Brownlee, P. M., Meisenberg, C. & Downs, J. A. The SWI/SNF chromatin remodelling complex: Its role in maintaining genome stability and preventing tumourigenesis. *DNA Repair (Amst)*. **32**, 127–133 (2015).
400. Dunaief, J. L. *et al.* The retinoblastoma protein and BRG1 form a complex and cooperate to induce cell cycle arrest. *Cell* **79**, 119–130 (1994).
401. Wong, A. K. C. *et al.* BRG1, a Component of the SWI-SNF Complex, Is Mutated in Multiple Human Tumor Cell Lines. *Cancer Res.* **60**, 6171–6177 (2000).
402. Kadoch, C. & Crabtree, G. R. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci. Adv.* **1**, e1500447 (2015).
403. Thompson, M. Polybromo-1: The chromatin targeting subunit of the PBAF complex. *Biochimie* **91**, 309–319 (2009).
404. Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* **46**, 225–233 (2014).
405. Brownlee, P. M., Chambers, A. L., Cloney, R., Bianchi, A. & Downs, J. A. BAF180 Promotes Cohesion and Prevents Genome Instability and Aneuploidy. *Cell Rep.* **6**, 973–981 (2014).
406. Niimi, A., Chambers, A. L., Downs, J. A. & Lehmann, A. R. A role for chromatin remodellers in replication of damaged DNA. *Nucleic Acids Res.* **40**, 7393–7403 (2012).
407. Niimi, A., Hopkins, S. R., Downs, J. A. & Masutani, C. The BAH domain of BAF180 is required for PCNA ubiquitination. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **779**, 16–23 (2015).
408. Vrouwe, M. G. *et al.* Increased DNA damage sensitivity of Cornelia de Lange syndrome cells: Evidence for impaired recombinational repair. *Hum. Mol. Genet.* **16**, 1478–1487 (2007).
409. Tatham, M. H. *et al.* RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. *Nat. Cell Biol.* **10**, 538–546 (2008).
410. Geoffroy, M.-C., Jaffray, E. G., Walker, K. J. & Hay, R. T. Arsenic-Induced SUMO-Dependent Recruitment of RNF4 into PML Nuclear Bodies. *Mol. Biol. Cell* **21**, 4227–4239 (2010).

411. Guo, B. & Sharrocks, A. D. Extracellular signal-regulated kinase mitogen-activated protein kinase signaling initiates a dynamic interplay between sumoylation and ubiquitination to regulate the activity of the transcriptional activator PEA3. *Mol. Cell. Biol.* **29**, 3204–3218 (2009).
412. Mukhopadhyay, D., Arnaoutov, A. & Dasso, M. The SUMO protease SENP6 is essential for inner kinetochore assembly. *J. Cell Biol.* **188**, 681–692 (2010).
413. van Hagen, M., Overmeer, R. M., Abolvardi, S. S. & Vertegaal, A. C. O. RNF4 and VHL regulate the proteasomal degradation of SUMO-conjugated Hypoxia-Inducible Factor-2 $\alpha$ . *Nucleic Acids Res.* **38**, 1922–1931 (2010).
414. Martin, N. *et al.* PARP-1 transcriptional activity is regulated by sumoylation upon heat shock. *EMBO J.* **28**, 3534–3548 (2009).
415. Kaiser, F. J., Möröy, T., Chang, G. T. G., Horsthemke, B. & Lüdecke, H.-J. The RING finger protein RNF4, a co-regulator of transcription, interacts with the TRPS1 transcription factor. *J. Biol. Chem.* **278**, 38780–38785 (2003).
416. Pfeiffer, A. *et al.* Ataxin-3 consolidates the MDC1-dependent DNA double-strand break response by counteracting the SUMO-targeted ubiquitin ligase RNF4. *EMBO J.* **36**, 1066–1083 (2017).
417. Galanty, Y., Belotserkovskaya, R., Coates, J. & Jackson, S. P. RNF4, a SUMO-targeted ubiquitin E3 ligase, promotes DNA double-strand break repair. *Genes Dev.* **26**, 1179–1195 (2012).
418. Luo, K. *et al.* CDK-mediated RNF4 phosphorylation regulates homologous recombination in S-phase. *Nucleic Acids Res.* **43**, 5465–5475 (2015).
419. Kuo, C.-Y., Li, X., Stark, J. M., Shih, H.-M. & Ann, D. K. RNF4 regulates DNA double-strand break repair in a cell cycle-dependent manner. *Cell Cycle* **15**, 787–798 (2016).
420. Xie, J. *et al.* RNF4-mediated polyubiquitination regulates the Fanconi anemia/BRCA pathway. *J. Clin. Invest.* **125**, 1523–1532 (2015).
421. Munakata, T., Adachi, N., Yokoyama, N., Kuzuhara, T. & Horikoshi, M. A human homologue of yeast anti-silencing factor has histone chaperone activity. *Genes to Cells* **5**, 221–233 (2000).
422. Tagami, H., Ray-Gallet, D., Almouzni, G. & Nakatani, Y. Histone H3.1 and H3.3 Complexes Mediate Nucleosome Assembly Pathways Dependent or Independent of DNA Synthesis. *Cell* **116**, 51–61 (2004).
423. Mello, J. A. *et al.* Human Asf1 and CAF-1 interact and synergize in a repair-coupled nucleosome assembly pathway. *EMBO Rep.* **3**, 329–334 (2002).
424. Tang, Y. *et al.* Structure of a human ASF1a/HIRA complex and insights into specificity of histone chaperone complex assembly. *Nat. Struct. Mol. Biol.* **13**, 921–929 (2006).
425. Mousson, F., Ochsenbein, F. & Mann, C. The histone chaperone Asf1 at the crossroads of chromatin and DNA checkpoint pathways. *Chromosoma* **116**, 79–93 (2007).
426. Zhang, R. *et al.* Formation of macroH2A-containing senescence-associated heterochromatin foci and senescence driven by ASF1a and HIRA. *Dev. Cell* **8**, 19–30 (2005).
427. Xie, W. *et al.* Histone H3 Lysine 56 Acetylation Is Linked to the Core Transcriptional Network in Human Embryonic Stem Cells. *Mol. Cell* **33**, 417–427 (2009).
428. Gonzalez-Munoz, E., Arboleda-Estudillo, Y., Otu, H. H. & Cibelli, J. B. Cell

- reprogramming: Histone chaperone ASF1A is required for maintenance of pluripotency and cellular reprogramming. *Science* (80-. ). **345**, 822–825 (2014).
429. Shi, H. *et al.* Roles of p53 and ASF1A in the Reprogramming of Sheep Kidney Cells to Pluripotent Cells. *Cell. Reprogram.* **17**, 441–452 (2015).
  430. Blanko, E. R., Kadyrova, L. Y. & Kadyrov, F. A. DNA mismatch repair interacts with CAF-1- and ASF1A-H3-H4-dependent histone (H3-H4)<sub>2</sub> tetramer deposition. *J. Biol. Chem.* **291**, 9203–9217 (2016).
  431. Hebbes, T. R., Thorne, A. W. & Crane-Robinson, C. A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J.* **7**, 1395–1402 (1988).
  432. Gross, D. S. & Garrard, W. T. Nuclease Hypersensitive Sites in Chromatin. *Annu. Rev. Biochem.* **57**, 159–197 (1988).
  433. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128**, 693–705 (2007).
  434. Swygert, S. G. & Peterson, C. L. Chromatin dynamics: Interplay between remodeling enzymes and histone modifications. *Biochim. Biophys. Acta* **1839**, 728–736 (2014).
  435. Riss, A. *et al.* Subunits of ADA-two-A-containing (ATAC) or Spt-Ada-Gcn5-acetyltransferase (SAGA) coactivator complexes enhance the acetyltransferase activity of GCN5. *J. Biol. Chem.* **290**, 28997–29009 (2015).
  436. Wang, L. & Y. R. Dent, S. Functions of SAGA in development and disease. *Epigenomics* **6**, 329–339 (2014).
  437. Nagy, Z. & Tora, L. Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. *Oncogene* **26**, 5341–5357 (2007).
  438. Spedale, G., Timmers, H. T. M. & Pijnappel, W. W. M. P. ATAC-king the complexity of SAGA during evolution. *Genes Dev.* **26**, 527–541 (2012).
  439. Feller, C., Forné, I., Imhof, A. & Becker, P. B. Global and specific responses of the histone acetylome to systematic perturbation. *Mol. Cell* **57**, 559–571 (2015).
  440. Suganuma, T. *et al.* ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding. *Nat. Struct. Mol. Biol.* **15**, 364–372 (2008).
  441. Lang, G. *et al.* The Tightly Controlled Deubiquitination Activity of the Human SAGA Complex Differentially Modifies Distinct Gene Regulatory Elements. *Mol. Cell. Biol.* **31**, 3734–3744 (2011).
  442. Bonnet, J. *et al.* The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. *Genes Dev.* **28**, 1999–2012 (2014).
  443. Guelman, S. *et al.* The Double-Histone-Acetyltransferase Complex ATAC Is Essential for Mammalian Development. *Mol. Cell. Biol.* **29**, 1176–1188 (2009).
  444. Col, E. *et al.* The Histone Acetyltransferase, hGCN5, Interacts with and Acetylates the HIV Transactivator, Tat. *J. Biol. Chem.* **276**, 28179–28184 (2001).
  445. Terreni, M. *et al.* GCN5-dependent acetylation of HIV-1 integrase enhances viral integration. *Retrovirology* **7**, 18 (2010).
  446. Zeng, L. & Zhou, M.-M. Bromodomain: An acetyl-lysine binding domain. *FEBS Lett.* **513**, 124–128 (2002).
  447. Martinez, E. *et al.* Human STAGA Complex Is a Chromatin-Acetylating Transcription Coactivator That Interacts with Pre-mRNA Splicing and DNA Damage-Binding Factors In

- Vivo. *Mol. Cell. Biol.* **21**, 6782–6795 (2001).
448. Pollard, K. J. & Peterson, C. L. Role for ADA/GCN5 products in antagonizing chromatin-mediated transcriptional repression. *Mol. Cell. Biol.* **17**, 6212–6222 (1997).
  449. Cosma, M. P., Tanaka, T. & Nasmyth, K. Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. *Cell* **97**, 299–311 (1999).
  450. Natarajan, K., Jackson, B. M., Zhou, H., Winston, F. & Hinnebusch, A. G. Transcriptional activation by Gcn4p involves independent interactions with the SWI/SNF complex and the SRB/mediator. *Mol. Cell* **4**, 657–664 (1999).
  451. Salma, N., Xiao, H., Mueller, E. & Imbalzano, A. N. Temporal Recruitment of Transcription Factors and SWI/SNF Chromatin-Remodeling Enzymes during Adipogenic Induction of the Peroxisome Proliferator-Activated Receptor  $\gamma$  Nuclear Hormone Receptor. *Mol. Cell. Biol.* **24**, 4651–4663 (2004).
  452. de la Serna, I. L. *et al.* MyoD targets chromatin remodeling complexes to the myogenin locus prior to forming a stable DNA-bound complex. *Mol. Cell. Biol.* **25**, 3997–4009 (2005).
  453. Agalioti, T., Chen, G. & Thanos, D. Deciphering the transcriptional histone acetylation code for a human gene. *Cell* **111**, 381–392 (2002).
  454. Jacobson, R. H., Ladurner, A. G., King, D. S. & Tijan, R. Structure and Function of a Human TAFII250 Double Bromodomain Module. *Science* (80-. ). **288**, 1422–1425 (2000).
  455. Kanno, T. *et al.* Selective Recognition of Acetylated Histones by Bromodomain Proteins Visualized in Living Cells. *Mol. Cell* **13**, 33–43 (2004).
  456. Josling, G. A., Selvarajah, S. A., Petter, M. & Duffy, M. F. The role of bromodomain proteins in regulating gene expression. *Genes (Basel)*. **3**, 320–343 (2012).
  457. Hassan, A. H., Neely, K. E. & Workman, J. L. Histone acetyltransferase complexes stabilize SWI/SNF binding to promoter nucleosomes. *Cell* **104**, 817–827 (2001).
  458. Hassan, A. H. *et al.* Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell* **111**, 369–379 (2002).
  459. Ferreira, M. E., Prochasson, P., Berndt, K. D., Workman, J. L. & Wright, A. P. H. Activator-binding domains of the SWI-SNF chromatin remodeling complex characterized in vitro are required for its recruitment to promoters in vivo. *FEBS J.* **276**, 2557–2565 (2009).
  460. Kastan, M. *et al.* Tandem bromodomains in the chromatin remodeler RSC recognize acetylated histone H3 Lys14. *EMBO J.* **23**, 1348–1359 (2004).
  461. VanDemark, A. P. *et al.* Autoregulation of the Rsc4 Tandem Bromodomain by Gcn5 Acetylation. *Mol. Cell* **27**, 817–828 (2007).
  462. Boyer, L. A. *et al.* Functional delineation of three groups of the ATP-dependent family of chromatin remodeling enzymes. *J. Biol. Chem.* **275**, 18864–18870 (2000).
  463. Park, J.-H. *et al.* Mammalian SWI/SNF complexes facilitate DNA double-strand break repair by promoting  $\gamma$ -H2AX induction. *EMBO J.* **25**, 3986–3997 (2006).
  464. Bhaumik, S. R. Distinct regulatory mechanisms of eukaryotic transcriptional activation by SAGA and TFIID. *Biochim. Biophys. Acta* **1809**, 97–108 (2011).
  465. Phillips, D. M. P. The presence of acetyl groups in histones. *Biochem. J.* **87**, 258–263 (1963).

466. Allfrey, V. G., Faulkner, R. & Mirsky, A. E. Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **51**, 786–794 (1964).
467. Verdin, E. & Ott, M. 50 Years of Protein Acetylation: From Gene Regulation To Epigenetics, Metabolism and Beyond. *Nat. Rev. Mol. Cell Biol.* **16**, 258–264 (2015).
468. Vidali, G., Boffa, L. C., Bradbury, E. M. & Allfrey, V. G. Butyrate suppression of histone deacetylation leads to accumulation of multiacetylated forms of histones H3 and H4 and increased DNase I sensitivity of the associated DNA sequences. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 2239–2243 (1978).
469. Candido, E. P. M., Reeves, R. & Davie, J. R. Sodium butyrate inhibits histone deacetylation in cultured cells. *Cell* **14**, 105–113 (1978).
470. Brownell, J. E. *et al.* Tetrahymena histone acetyltransferase A: A homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* **84**, 843–851 (1996).
471. Taunton, J., Hassig, C. A. & Schreiber, S. L. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science (80-. )*. **272**, 408–411 (1996).
472. Mizzen, C. A. *et al.* The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. *Cell* **87**, 1261–1270 (1996).
473. Ogryzko, V. V., Schiltz, R. L., Russanova, V., Howard, B. H. & Nakatani, Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* **87**, 953–959 (1996).
474. Wapenaar, H. & Dekker, F. J. Histone acetyltransferases: challenges in targeting bi-substrate enzymes. *Clin. Epigenetics* **8**, 59 (2016).
475. Dutnall, R. N., Tafrov, S. T., Sternglanz, R. & Ramakrishnan, V. Structure of the histone acetyltransferase Hat1: a paradigm for the GCN5-related N-acetyltransferase superfamily. *Cell* **94**, 427–438 (1998).
476. Trievel, R. C. *et al.* Crystal structure and mechanism of histone acetylation of the yeast GCN5 transcriptional coactivator. *Proc. Natl. Acad. Sci.* **96**, 8931–8936 (1999).
477. Clements, A. *et al.* Crystal structure of the histone acetyltransferase domain of the human PCAF transcriptional regulator bound to coenzyme A. *EMBO J.* **18**, 3521–3532 (1999).
478. Tanner, K. G. *et al.* Catalytic Mechanism and Function of Invariant Glutamic Acid 173 from the Histone Acetyltransferase GCN5 Transcriptional Coactivator. *J. Biol. Chem.* **274**, 18157–18160 (1999).
479. Tanner, K. G., Langer, M. R., Kim, Y. & Denu, J. M. Kinetic mechanism of the histone acetyltransferase GCN5 from yeast. *J. Biol. Chem.* **275**, 22048–22055 (2000).
480. Rojas, J. R. *et al.* Structure of Tetrahymena GCN5 bound to coenzyme A and a histone H3 peptide. *Nature* **401**, 93–98 (1999).
481. L'Hernault, S. W. & Rosenbaum, J. L. Chlamydomonas  $\alpha$ -tubulin is posttranslationally modified by acetylation on the  $\epsilon$ -amino group of a lysine. *Biochemistry* **24**, 473–478 (1985).
482. Gu, W. & Roeder, R. G. Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell* **90**, 595–606 (1997).
483. Kiernan, R. E. *et al.* HIV-1 Tat transcriptional activity is regulated by acetylation. *EMBO J.* **18**, 6106–6118 (1999).

484. Ott, M. *et al.* Acetylation of the HIV-1 tat protein by p300 is important for its transcriptional activity. *Curr. Biol.* **9**, 1489–1492 (1999).
485. Chen, L.-F., Fischle, W., Verdin, E. & Greene, W. C. Duration of Nuclear NF- $\kappa$ B Action Regulated by Reversible Acetylation. *Science (80-. )*. **293**, 1653–1657 (2001).
486. Kim, S. C. *et al.* Substrate and Functional Diversity of Lysine Acetylation Revealed by a Proteomics Survey. *Mol. Cell* **23**, 607–618 (2006).
487. Choudhary, C. *et al.* Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science (80-. )*. **325**, 834–840 (2009).
488. Guha, M. HDAC inhibitors still need a home run, despite recent approval. *Nat. Rev. Drug Discov.* **14**, 225–226 (2015).
489. Tanaka, Y. *et al.* Extensive brain hemorrhage and embryonic lethality in a mouse null mutant of CREB-binding protein. *Mech. Dev.* **95**, 133–145 (2000).
490. Gupta, A. *et al.* The Mammalian Ortholog of Drosophila MOF That Acetylates Histone H4 Lysine 16 Is Essential for Embryogenesis and Oncogenesis. *Mol. Cell. Biol.* **28**, 397–409 (2008).
491. Furdas, S. D., Kannan, S., Sippl, W. & Jung, M. Small molecule inhibitors of histone acetyltransferases as epigenetic tools and drug candidates. *Arch. Pharm. (Weinheim)*. **345**, 7–21 (2012).
492. Lau, O. D. *et al.* HATs off: selective synthetic inhibitors of the histone acetyltransferases p300 and PCAF. *Mol. Cell* **5**, 589–595 (2000).
493. Balasubramanyam, K., Swaminathan, V., Ranganathan, A. & Kundu, T. K. Small molecule modulators of histone acetyltransferase p300. *J. Biol. Chem.* **278**, 19134–19140 (2003).
494. Hemshekhar, M., Santhosh, M. S., Kemparaju, K. & Girish, K. S. Emerging Roles of Anacardic Acid and Its Derivatives: A Pharmacological Overview. *Basic Clin. Pharmacol. Toxicol.* **110**, 122–132 (2012).
495. Balasubramanyam, K. *et al.* Curcumin, a novel p300/CREB-binding protein-specific inhibitor of acetyltransferase, represses the acetylation of histone/nonhistone proteins and histone acetyltransferase-dependent chromatin transcription. *J. Biol. Chem.* **279**, 51163–51171 (2004).
496. Balasubramanyam, K. *et al.* Polyisoprenylated benzophenone, garcinol, a natural histone acetyltransferase inhibitor, represses chromatin transcription and alters global gene expression. *J. Biol. Chem.* **279**, 33716–33726 (2004).
497. Jeffers, V. *et al.* Garcinol inhibits GCN5-mediated lysine acetyltransferase activity and prevents replication of the parasite *Toxoplasma gondii*. *Antimicrob. Agents Chemother.* **60**, 2164–2170 (2016).
498. Modak, R. *et al.* Probing p300/CBP associated factor (PCAF)-dependent pathways with a small molecule inhibitor. *ACS Chem. Biol.* **8**, 1311–1323 (2013).
499. Ghizzoni, M., Boltjes, A., Graaf, C. De, Haisma, H. J. & Dekker, F. J. Improved inhibition of the histone acetyltransferase PCAF by an anacardic acid derivative. *Bioorg. Med. Chem.* **18**, 5826–5834 (2010).
500. Costi, R. *et al.* Cinnamoyl compounds as simple molecules that inhibit p300 histone acetyltransferase. *J. Med. Chem.* **50**, 1973–1977 (2007).
501. Arif, M. *et al.* Nitric Oxide-Mediated Histone Hyperacetylation in Oral Cancer: Target for a Water-Soluble HAT Inhibitor, CTK7A. *Chem. Biol.* **17**, 903–913 (2010).

502. Biel, M., Kretsovali, A., Karatzali, E., Papamatheakis, J. & Giannis, A. Design, synthesis, and biological evaluation of a small-molecule inhibitor of the histone acetyltransferase Gcn5. *Angew. Chemie - Int. Ed.* **43**, 3974–3976 (2004).
503. Aquea, F., Timmermann, T. & Herrera-Vásquez, A. Chemical inhibition of the histone acetyltransferase activity in *Arabidopsis thaliana*. *Biochem. Biophys. Res. Commun.* **483**, 664–668 (2017).
504. Mai, A. *et al.* Small-molecule inhibitors of histone acetyltransferase activity: Identification and biological properties. *J. Med. Chem.* **49**, 6897–6907 (2006).
505. Stimson, L. *et al.* Isothiazolones as inhibitors of PCAF and p300 histone acetyltransferase activity. *Mol. Cancer Ther.* **4**, 1521–1532 (2005).
506. Ghizzoni, M., Haisma, H. J. & Dekker, F. J. Reactivity of isothiazolones and isothiazolone-1-oxides in the inhibition of the PCAF histone acetyltransferase. *Eur. J. Med. Chem.* **44**, 4855–4861 (2009).
507. Dekker, F. J., Ghizzoni, M., Meer, N. van der, Wisastra, R. & Haisma, H. J. Inhibition of the PCAF histone acetyl transferase and cell proliferation by isothiazolones. *Bioorg. Med. Chem.* **17**, 460–466 (2009).
508. Gorsuch, S. *et al.* Synthesis of isothiazol-3-one derivatives as inhibitors of histone acetyltransferases (HATs). *Bioorg. Med. Chem.* **17**, 467–474 (2009).
509. Furdas, S. D. *et al.* Synthesis and biological testing of novel pyridoisothiazolones as histone acetyltransferase inhibitors. *Bioorg. Med. Chem.* **19**, 3678–3689 (2011).
510. Wisastra, R. *et al.* Isothiazolones; thiol-reactive inhibitors of cysteine protease cathepsin B and histone acetyltransferase PCAF. *Org. Biomol. Chem.* **9**, 1817–1822 (2011).
511. Chimenti, F. *et al.* A novel histone acetyltransferase inhibitor modulating Gcn5 network: cyclopentylidene-[4-(4'-chlorophenyl)thiazol-2-yl]hydrazone. *J. Med. Chem.* **52**, 530–536 (2009).
512. Trisciuglio, D. *et al.* CPTH6, a thiazole derivative, induces histone hypoacetylation and apoptosis in human leukemia cells. *Clin. Cancer Res.* **18**, 475–486 (2012).
513. Secci, D. *et al.* Synthesis of a novel series of thiazole-based histone acetyltransferase inhibitors. *Bioorg. Med. Chem.* **22**, 1680–1689 (2014).
514. Carradori, S. *et al.* Evaluation of a large library of (thiazol-2-yl)hydrazones and analogues as histone acetyltransferase inhibitors: Enzyme and cellular studies. *Eur. J. Med. Chem.* **80**, 569–578 (2014).
515. Martile, M. Di *et al.* Histone acetyltransferase inhibitor CPTH6 preferentially targets lung cancer stem-like cells. *Oncotarget* **7**, 11332–11348 (2016).
516. Kumar, A. *et al.* Designing novel inhibitors against histone acetyltransferase (HAT: GCN5) of *Plasmodium falciparum*. *Eur. J. Med. Chem.* **138**, 26–37 (2017).
517. Baell, J. & Walters, M. A. Chemical con artists foil drug discovery. *Nature* **513**, 481–483 (2014).
518. Grant, P. A. *et al.* Expanded lysine acetylation specificity of Gcn5 in native complexes. *J. Biol. Chem.* **274**, 5895–5900 (1999).
519. Bowers, E. M. *et al.* Virtual ligand screening of the p300/CBP histone acetyltransferase: Identification of a selective small molecule inhibitor. *Chem. Biol.* **17**, 471–482 (2010).
520. Yan, G. *et al.* Selective Inhibition of p300 HAT Blocks Cell Cycle Progression, Induces



- Cellular Senescence, and Inhibits the DNA Damage Response in Melanoma Cells. *J. Invest. Dermatol.* **133**, 2444–2452 (2013).
521. Santer, F. R. *et al.* Inhibition of the acetyltransferases p300 and CBP reveals a targetable function for p300 in the survival and invasion pathways of prostate cancer cell lines. *Mol. Cancer Ther.* **10**, 1644–1655 (2011).
  522. Fujisawa, T. & Filippakopoulos, P. Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. *Nat. Rev. Mol. Cell Biol.* **18**, 246–262 (2017).
  523. Glozak, M. A. & Seto, E. Histone deacetylases and cancer. *Oncogene* **26**, 5420–5432 (2007).
  524. Shortt, J., Ott, C. J., Johnstone, R. W. & Bradner, J. E. A chemical probe toolbox for dissecting the cancer epigenome. *Nat. Rev. Cancer* **17**, 160–183 (2017).
  525. Tamkun, J. W. *et al.* brhma: A regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2 SWI2. *Cell* **68**, 561–572 (1992).
  526. Haynes, S. R. *et al.* The bromodomain: A conserved sequence found in human, Drosophila and yeast proteins. *Nucleic Acids Res.* **20**, 2603 (1992).
  527. Daguer, J.-P. *et al.* Identification of covalent bromodomain binders through DNA display of small molecules. *Angew. Chemie - Int. Ed.* **54**, 6057–6061 (2015).
  528. Dancy, B. M. & Cole, P. A. Protein lysine acetylation by p300/CBP. *Chem. Rev.* **115**, 2419–2452 (2015).
  529. Byrd, K. N. & Shearn, A. ASH1, a Drosophila trithorax group protein, is required for methylation of lysine 4 residues on histone H3. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 11535–11540 (2003).
  530. Ali, M., Hom, R. A., Blakeslee, W., Ikenouye, L. & Kutateladze, T. G. Diverse functions of PHD fingers of the MLL/KMT2 subfamily. *Biochim. Biophys. Acta* **1843**, 366–371 (2014).
  531. Rahman, S. *et al.* The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3. *Mol. Cell. Biol.* **31**, 2641–2652 (2011).
  532. Jacobson, R. H., Ladurner, A. G., King, D. S. & Tijan, R. Structure and Function of a Human TAFII250 Double Bromodomain Module. *Science (80-. )*. **288**, 1422–1425 (2000).
  533. Wassarman, D. A. & Sauer, F. TAFII250: a transcription toolbox. *J. Cell Sci.* **114**, 2895–2920 (2001).
  534. Tsai, W.-W. *et al.* TRIM24 links a non-canonical histone signature to breast cancer. *Nature* **468**, 927–932 (2010).
  535. Cammas, F. *et al.* Cell differentiation induces TIF1 $\beta$  association with centromeric heterochromatin via an HP1 interaction. *J. Cell Sci.* **115**, 3439–3448 (2002).
  536. Ivanov, A. V. *et al.* PHD Domain-Mediated E3 Ligase Activity Directs Intramolecular Sumoylation of an Adjacent Bromodomain which is Required for Gene Silencing. *Mol. Cell* **28**, 823–837 (2007).
  537. Dupont, S. *et al.* FAM/USP9x, a Deubiquitinating Enzyme Essential for TGF $\beta$  Signaling, Controls Smad4 Monoubiquitination. *Cell* **136**, 123–135 (2009).
  538. Huether, R. *et al.* The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. *Nat. Commun.* **5**, 3630 (2014).
  539. Filippakopoulos, P. & Knapp, S. Targeting bromodomains: epigenetic readers of lysine

- acetylation. *Nat. Rev. Drug Discov.* **13**, 337–356 (2014).
540. Jeanmougin, F., Wurtz, J.-M., Douarin, B. Le, Chambon, P. & Losson, R. The bromodomain revisited. *Trends Biochem. Sci.* **22**, 151–153 (1997).
  541. Dhalluin, C. *et al.* Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**, 491–496 (1999).
  542. Mujtaba, S. *et al.* Structural Basis of Lysine-Acetylated HIV-1 Tat Recognition by PCAF Bromodomain. *Mol. Cell* **9**, 575–586 (2002).
  543. Mujtaba, S. *et al.* Structural Mechanism of the Bromodomain of the Coactivator CBP in p53 Transcriptional Activation. *Mol. Cell* **13**, 251–263 (2004).
  544. Nady, N., Min, J., Kareta, M. S., Chédin, F. & Arrowsmith, C. H. A SPOT on the chromatin landscape? Histone peptide arrays as a tool for epigenetic research. *Trends Biochem. Sci.* **33**, 305–313 (2008).
  545. Ferri, E., Petosa, C. & McKenna, C. E. Bromodomains: Structure, function and pharmacology of inhibition. *Biochem. Pharmacol.* **106**, 1–18 (2016).
  546. Schröder, S. *et al.* Two-pronged binding with bromodomain-containing protein 4 liberates positive transcription elongation factor b from inactive ribonucleoprotein complexes. *J. Biol. Chem.* **287**, 1090–1099 (2012).
  547. Morinière, J. *et al.* Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature* **461**, 664–668 (2009).
  548. Zeng, L. *et al.* Selective small molecules blocking HIV-1 Tat and coactivator PCAF association. *J. Am. Chem. Soc.* **127**, 2376–2377 (2005).
  549. Pan, C. *et al.* Structure-guided optimization of small molecules inhibiting human immunodeficiency virus 1 Tat association with the human coactivator p300/CREB binding protein-associated factor. *J. Med. Chem.* **50**, 2285–2288 (2007).
  550. Sachchidanand *et al.* Target structure-based discovery of small molecules that block human p53 and CREB binding protein association. *Chem. Biol.* **13**, 81–90 (2006).
  551. Mirguet, O. *et al.* Discovery of epigenetic regulator I-BET762: Lead optimization to afford a clinical candidate inhibitor of the BET bromodomains. *J. Med. Chem.* **56**, 7501–7515 (2013).
  552. Vidler, L. R., Brown, N., Knapp, S. & Hoelder, S. Druggability Analysis and Structural Classification of Bromodomain Acetyl-lysine Binding Sites. *J. Med. Chem.* **55**, 7346–7359 (2012).
  553. Gallenkamp, D., Gelato, K. A., Haendler, B. & Weinmann, H. Bromodomains and their pharmacological inhibitors. *ChemMedChem* **9**, 438–464 (2014).
  554. Noel, J. K. *et al.* Development of the BET bromodomain inhibitor OTX015. *Mol. Cancer Ther.* **12**, C244 (2013).
  555. Boi, M. *et al.* The BET bromodomain inhibitor OTX015 affects pathogenetic pathways in preclinical B-cell tumor models and synergizes with targeted drugs. *Clin. Cancer Res.* **21**, 1628–1638 (2015).
  556. Coudé, M.-M. *et al.* BET inhibitor OTX015 targets BRD2 and BRD4 and decreases c-MYC in acute leukemia cells. *Oncotarget* **6**, 17698–17712 (2015).
  557. Hewings, D. S. *et al.* 3,5-Dimethylisoxazoles Act As Acetyl-Lysine-Mimetic Bromodomain Ligands. *J. Med. Chem.* **54**, 6761–6770 (2011).

558. Philpott, M. *et al.* Bromodomain-peptide displacement assays for interactome mapping and inhibitor discovery. *Mol. Biosyst.* **7**, 2899–2908 (2011).
559. Dawson, M. A. *et al.* Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature* **478**, 529–533 (2011).
560. Fish, P. V. *et al.* Identification of a chemical probe for bromo and extra C-terminal bromodomain inhibition through optimization of a fragment-derived hit. *J. Med. Chem.* **55**, 9831–9837 (2012).
561. Picaud, S. *et al.* PFI-1, a highly selective protein interaction inhibitor, targeting BET bromodomains. *Cancer Res.* **73**, 3336–3346 (2013).
562. Picaud, S. *et al.* RVX-208, an inhibitor of BET transcriptional regulators with selectivity for the second bromodomain. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 19754–19759 (2013).
563. Borah, J. C. *et al.* A small molecule binding to the coactivator CREB-binding protein blocks apoptosis in cardiomyocytes. *Chem. Biol.* **18**, 531–541 (2011).
564. Hay, D. A. *et al.* Discovery and Optimization of Small-Molecule Ligands for the CBP/p300 Bromodomains. *J. Am. Chem. Soc.* **136**, 9308–9319 (2014).
565. Drouin, L. *et al.* Structure enabled design of BAZ2-ICR, a chemical probe targeting the bromodomains of BAZ2A and BAZ2B. *J. Med. Chem.* **58**, 2553–2559 (2015).
566. Chen, P. *et al.* Discovery and Characterization of GSK2801, a Selective Chemical Probe for the Bromodomains BAZ2A and BAZ2B. *J. Med. Chem.* **59**, 1410–1424 (2016).
567. Palmer, W. S. *et al.* Structure-Guided Design of IACS-9571, a Selective High-Affinity Dual TRIM24-BRPF1 Bromodomain Inhibitor. *J. Med. Chem.* **59**, 1440–1454 (2016).
568. Clark, P. G. K. *et al.* LP99: Discovery and synthesis of the first selective BRD7/9 bromodomain inhibitor. *Angew. Chemie Int. Ed.* **54**, 6217–6221 (2015).
569. Xu, Y. & Vakoc, C. R. Targeting Cancer Cells with BET Bromodomain Inhibitors. *Cold Spring Harb. Perspect. Med.* **7**, a026674 (2017).
570. Duffy, B. C. *et al.* Discovery of a new chemical series of BRD4(1) inhibitors using protein-ligand docking and structure-guided design. *Bioorganic Med. Chem. Lett.* **25**, 2818–2823 (2015).
571. Crawford, T. D. *et al.* Discovery of a Potent and Selective in Vivo Probe (GNE-272) for the Bromodomains of CBP/EP300. *J. Med. Chem.* **59**, 10549–10563 (2016).
572. Ward, W. H. J. & Holdgate, G. A. in *Progress in Medicinal Chemistry* (eds. King, F. D. & Oxford, A. W.) **38**, 309–376 (Elsevier Science B.V., 2001).
573. Crawford, T. D. *et al.* Diving into the Water: Inducible Binding Conformations for BRD4, TAF1(2), BRD9, and CECR2 Bromodomains. *J. Med. Chem.* **59**, 5391–5402 (2016).
574. Hudson, B. P., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. Solution structure and acetyl-lysine binding activity of the GCN5 bromodomain. *J. Mol. Biol.* **304**, 355–370 (2000).
575. Hu, P. *et al.* Fluorescence Polarization for the Evaluation of Small-Molecule Inhibitors of PCAF BRD/Tat-Ack50 Association. *ChemMedChem* **9**, 928–931 (2014).
576. Montgomery, D. C., Sorum, A. W. & Meier, J. L. Chemoproteomic profiling of lysine acetyltransferases highlights an expanded landscape of catalytic acetylation. *J. Am. Chem. Soc.* **136**, 8669–8676 (2014).
577. Navratilova, I. *et al.* Discovery of New Bromodomain Scaffolds by Biosensor Fragment

- Screening. *ACS Med. Chem. Lett.* (2016). doi:10.1021/acsmedchemlett.6b00154
578. Albrecht, B. K. *et al.* Phthalazine derivatives of formula (I) as PCAF and GCN5 inhibitors for use in the treatment of cancer. WO 2016/036954 A1. (2016).
579. Albrecht, B. K. *et al.* Therapeutic compounds and uses thereof. WO 2016/036873 A1. (2016).
580. Albrecht, B. K. *et al.* Pyridazinone derivatives and their use in the treatment of cancer. WO 2016/112298 A1. (2016).
581. Moustakim, M. *et al.* Discovery of a PCAF Bromodomain Chemical Probe. *Angew. Chemie - Int. Ed.* **56**, 827–831 (2017).
582. Humphreys, P. G. *et al.* Discovery of a potent, cell penetrant, and selective p300/CBP-associated factor (PCAF)/general control nonderepressible 5 (GCN5) bromodomain chemical probe. *J. Med. Chem.* **60**, 695–709 (2017).
583. Sakamoto, K. M. *et al.* Protacs: chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8554–8559 (2001).
584. Zengerle, M., Chan, K.-H. & Ciulli, A. Selective Small Molecule Induced Degradation of the BET Bromodomain Protein BRD4. *ACS Chem. Biol.* **10**, 1770–1777 (2015).
585. Raina, K. *et al.* PROTAC-induced BET protein degradation as a therapy for castration-resistant prostate cancer. *Proc. Natl. Acad. Sci.* **113**, 7124–7129 (2016).
586. Winter, G. E. *et al.* Phthalimide conjugation as a strategy for in vivo target protein degradation. *Science (80-. )*. **348**, 1376–1381 (2015).
587. Lu, J. *et al.* Hijacking the E3 Ubiquitin Ligase Cereblon to Efficiently Target BRD4. *Chem. Biol.* **22**, 755–763 (2015).
588. Ran, X. *et al.* Structure-Based Design of  $\gamma$ -Carboline Analogues as Potent and Specific BET Bromodomain Inhibitors. *J. Med. Chem.* **58**, 4927–4939 (2015).
589. Zhou, B. *et al.* Discovery of a Small-Molecule Degradator of Bromodomain and Extra-Terminal (BET) Proteins with Picomolar Cellular Potencies and Capable of Achieving Tumor Regression. *J. Med. Chem.* [Epub ahead of print] (2017). doi:10.1021/acs.jmedchem.6b01816

## VII Appendix

### VII.1. SMC Multiple Sequence Alignments

```

SMC1A_HUMAN -----0
SMC1B_HUMAN -----0
SMC2_HUMAN -----0
SMC3_HUMAN -----0
SMC4_HUMAN -MPRKGTQPSTARRREEGPPPPSPDGASSDAEPEPPSGRTE SPATAAETASEELDNRSL EE-60
SMC5_HUMAN -----MATPSKKTSTPSPQPSKR---ALPRDPS-----SEVPSKR-----32
SMC6_HUMAN -----MAKR-KEENFSSPKNAK----RPRQEE-----LEDFDKDG D--31
RAD50_HUMAN -----0

SMC1A_HUMAN -----MGFLKLI E I E N F K S Y K ---GRQI I GP--FQ-RF T A I I GP-33
SMC1B_HUMAN -----MAHLELLLVEN F K S W R ---GRQV I GP--FR-RF T C I I GP-33
SMC2_HUMAN -----MH I K S I I L E G F K S Y A ---QRTE V NG--FDPL F N A I T G L-33
SMC3_HUMAN -----MY I K Q V I I Q G F R S Y R ---DQ T I V DP--FSSKH N V I V GR-33
SMC4_HUMAN -ILNSIPPPPPPPAMTNEAGAPRLMITHIVNQN F K S Y A ---GEKI L GP--FHKRF S C I I GP-114
SMC5_HUMAN -----KNSAPQLPLLLQSSGPFVEGSIVRISME N F L T Y D ---ICEVSPGPHL N M I V G A-81
SMC6_HUMAN -----EDECKGTTL---TAAEVGI I E S I H L K N F M C H S ---MLGPFKFGS N V N F V V G N-77
RAD50_HUMAN -----MSRIE K M S I L G V R S F G I E D K D K Q I I -T---FFSPL T I L V GP-37

SMC1A_HUMAN -NGSGKSN L M D A I S F V L G E K -TSNLRVKTLRDL I H G A P V G K P A A N R A F V ---S M V Y S E E G-88
SMC1B_HUMAN -NGSGKSN V M D A L S F V M G E K -I A N L R V K N I Q E L I H G A H I G K P I S S A S V ---K I I Y V E S-88
SMC2_HUMAN -NGSGKSN I L D S I C F L L G I S N L S Q V R A S N L Q D L V Y K N G Q A G I ---T K A S V ---S I T F D N S-87
SMC3_HUMAN -NGSGKSN F F Y A I Q F V L S D E -F S H L R P E Q R L A L L H E G T G P R V ---I S A F V ---E I I F D N S-86
SMC4_HUMAN -NGSGKSN V I D S M L F V F G Y R -A Q K I R S K K L S V L I H N S D E H K D I Q S C T V E V H F Q K I I -D K E G-172
SMC5_HUMAN -NGT G K S S I V C A I C L G L A G K P A F M G R A D K V G F F V K R G C S R G M ---V ---E I E L F R A S-131
SMC6_HUMAN -NGSGKSA V L T A L I V G L G G R A V A T N R G S S L K G F V K D G Q N S A D ---I ---S I T L R N R G-127
RAD50_HUMAN -N G A G K T T I I E C L K Y I C T G D -F P P G - -T K G N T F V H D P K V A Q E T D V R A Q I ---R L Q F R D V N-90

SMC1A_HUMAN -AEDRTF-----ARVIVGG-SSEYK I N N K V V Q L ---HEYSEE L E K L G-125
SMC1B_HUMAN -GEEKTF-----ARIIRGG-CSEFR F D N L V S R ---S V Y I A E L I G-125
SMC2_HUMAN -KKQSP L G F E ---V H D E I T V T R Q V V I G G -R N K Y L I N G V N A N N ---T R V Q D L F C S V G-135
SMC3_HUMAN -NR L P -----I D K E E V S L R R V I G A K -K D Q Y F L D K K M V T K ---N D V M N L L E S A G-129
SMC4_HUMAN -DDYEVI-----P N S N F Y V S R T A C R D N -T S V Y H I S G K K K T F ---K D V G N L I R S H G-217
SMC5_HUMAN -G-----N L V I T R E I D V A K N Q S F W F I N K K S T T Q ---K I V E E K V A A L N-169
SMC6_HUMAN -DDAFKA--S---V Y G N S I L I Q Q H I S I D G S R S Y K L K S A T G S V V S T R K E E L I A I L D H F N-179
RAD50_HUMAN -G E L I A V Q R S M V C T Q K S K K T E F K T L E G V I T R T -K H G E K V S L S S K C A ---E I D R E M I S S L G-145

SMC1A_HUMAN -I L I K A R N F L V F Q G A V E S I A M K N P K E R T A L ---FEEISRS G E L A Q E Y D K R K K E M V K A-178
SMC1B_HUMAN -I V K A Q N C L V F Q G T V E S I S V K K P K E R T Q F ---FEEISTSGELIGEYEEKKRKLQKA-178
SMC2_HUMAN -L N V N N P H F L I M Q G R I T K V L N M K P P E I L S M ---IEEAAGTRM---Y E Y K K I A A Q K T-184
SMC3_HUMAN -F S R S N P Y Y I V K Q G K I N M A T A P D S Q R L K L ---L R E V A G T R V ---Y D E R K E E S I S L-178
SMC4_HUMAN -I D L D H N R F L I L Q G E V E Q I A M M K P K G Q T E H D E G M L E Y L E D I I G C G R L N E P I K V L C R R V E I L-277
SMC5_HUMAN -I Q V G N L C Q F L P Q D K V G E F A K L S K ---I E L L -E A T E K ---S-202
SMC6_HUMAN -I Q V D N P V S V L T Q E M S K Q F L Q S K N E G D K ---Y K F F M K A T Q L ---216
RAD50_HUMAN -V S K A V L N N V I F C H Q E D S N W ---P L S E G K ---A L K Q K F D E I F S A T R Y I K A L E T L R Q ---V-195

SMC1A_HUMAN -E E D T Q F N Y H R K K ---N I A A E ---R---K E A K Q -E K E A D R Y Q R L K D E V V R A-219
SMC1B_HUMAN -E E D A Q F N F N K K K ---N I A A E ---R---R Q A K L -E K E A E R Y Q S L L E E L K M N-219
SMC2_HUMAN -I E K K E A K L K E I K ---T I L E E ---E I T ---P T I Q K -L K E E R S S Y L E Y Q K -V M R E-226
SMC3_HUMAN -M K E T E G K R E K I N E L L K Y I E ---R L ---H T L E E -E K E E L A Q Y Q K W D K -M R R A-222
SMC4_HUMAN -N E H R G E K L N R V K ---M V E K E ---K---D A L E G -E K N I A I E F L T L E N E I F R K-318
SMC5_HUMAN -I-----G P P E M H K Y H C E L K ---N L R E K E K Q L E T S C K E K T E Y L Q K M V Q R N E R Y-246
SMC6_HUMAN -E-----Q M K E D Y S Y I M E T K E R T K E Q I H Q G E E R L T E L K R Q C V E K E E R F Q S I A G ---263
RAD50_HUMAN -R Q T Q G Q K V K E Y Q M E L K Y L K Q Y K E K A C E I R D Q I T S K E A Q L T S -S K E I ---240

SMC1A_HUMAN -Q V Q L -Q L F K L Y H N E ---V E I E K L N K E ---L A S K N K E I E K D K K-254
SMC1B_HUMAN -K I Q L -Q L F Q L Y H N E ---K K I H L L N T K ---L E H V N R D L S V K R E-254
SMC2_HUMAN -I E H L S R L Y I A Y Q F L L A E D T -K V ---R S A E E L K E M Q D-258
SMC3_HUMAN -L E ---Y T I Y N Q E L N E T R A K L D E L S A K ---R E T S G E K S Q R L R D-258
SMC4_HUMAN -K N H V -C Q Y Y I Y E L Q ---K S N I L S N E M K A K N K D V ---K E K I H E D T K E I N E-353
SMC5_HUMAN -K Q D V E R F Y E R K R H L ---D L I E M L E A K R P W V E Y E N V R Q E Y E E V K L V R D R V K E E V R K L K E-301
SMC6_HUMAN -----L S T M ---K ---T N L E S L K H E M A W A V V N E I E K Q L N ---A I R D N I K I ---299
RAD50_HUMAN -----V K S Y E N E L D P L K N R L K I E I H N L S K I M ---K L D N E I K A L D S R K K M E K D N S E L E E-291

SMC1A_HUMAN -----R M D K V E D E L K E K K K E L ---G K M M R E Q Q Q I E K E ---I-284
SMC1B_HUMAN -----S L S H H E N I V K A R K K E H ---G M L T R Q L Q Q T E K E ---L-284
SMC2_HUMAN -----K V I K L Q E E L S E N D K K I ---K A L N H E I E E L E K R K D K E T G G I L-296
SMC3_HUMAN -----A Q Q D A R D K M E D I E R Q V ---R E L K T K I S A M K E E K E ---289
SMC4_HUMAN -----K S N I L S N E M K A K N K D V ---K D T E K K L N K I T F ---I-383
SMC5_HUMAN -G Q I P V -T C R I E E M E N E R H N L E A R I K E K A T D I K E A S Q K C K Q K Q D V I E R K D K H I E E L Q Q A L I-360
SMC6_HUMAN -G-----E D R A A R L D R K E E Q Q V R L N E A E Q K Y K D I Q D K L E K I S E E T N A R A P E C M-347
RAD50_HUMAN -K M E K V F Q G T D E Q L N D L Y H N H Q R T V R E K E R K L ---V D C H R E L E K L N K E S R ---L L-339

SMC1A_HUMAN -K E -K D S E L N Q K R P Q Y I K A K E N T S H K I K K L E A A K K S L Q N -----321
SMC1B_HUMAN -K S -V E T L N Q K R P Q Y I K A K E N T S H H L K K L D V A K K S I K D -----321
SMC2_HUMAN -R S -L E D A L A E A Q R V N T K -----S Q S A F D L K K K N L A C E E S K R K E L E K -----N M V E D S-342
SMC3_HUMAN -----Q L S A E R Q E Q I K -----Q R T K L E L K A K D L Q D E L A G N S E Q R K -----R L L K E R-330
SMC4_HUMAN -E E -N K E K F T Q L D L E D V Q V R E K L K H A T S K A K K L E K Q L Q K D K E K V E E F S I P A K S N N I N E T-442
SMC5_HUMAN -V K -Q N E E L D R -----V D V Q V R E K L K H A T S K A K K L E K Q L Q K D K E K V E E F S I P A K S N N I N E T-376
SMC6_HUMAN -A L -K A D V V A K -----K R A Y N E A-363
RAD50_HUMAN -N Q E K S E L L V E Q G R L Q L Q A D R H Q E H I R -A R D S L I Q S L A T Q L E -L D G F E R G P F S E -R Q I K N F-396

```

SMC1A\_HUMAN - - - - - A Q K H Y K K R - - - - - K G D M D E L E K E - - - - - M L S V E - 344  
SMC1B\_HUMAN - - - - - S E K Q C S K Q - - - - - E D D I K A L E T E - - - - - L A D L D - 344  
SMC2\_HUMAN - K T L A - - - - - A K E K V K K I - - - - - T D G L H A L Q E A S - - - - - N K D A E A L A - 374  
SMC3\_HUMAN - Q K L L E - K I E E K Q K E L A E T - - - - - E P K F N S V K E K E - - - - - E R G I A R L A - 366  
SMC4\_HUMAN - T T R N N A L E K E K E K E K L - - - - - K E V M D S L K Q E - - - - - T Q G L Q - 475  
SMC5\_HUMAN - R K M I E - - - - - D L Q N E L K T T E N C E N L Q P Q I D A I T N D L R R I Q D E K A L C E G E I I D K R R E R E T L E - 432  
SMC6\_HUMAN - E V L Y N - - - - - R S L N E Y K A - - - - - L K K D D E Q L C K R I - - - - - 388  
RAD50\_HUMAN - H K L V R E - - - - - R Q E G E A K T A - - - - - N Q L M N D F A E K E - - - - - T - - - - - L K - 425

SMC1A\_HUMAN - K A R - - - - - Q E F E E R M E E E S Q S Q G R D L T L E E N Q V K K Y - - - - - H R L K E E A S K R A A - 387  
SMC1B\_HUMAN - A A W - - - - - R S F E K Q I E E E I L H K K R D I E L E A S Q L D R Y - - - - - K E L K E Q V R K K V A - 387  
SMC2\_HUMAN - A A Q - - - - - Q H F N A - - - - - V S A G - - - - - L S S N E D G A E A T - - - - - L A - - - - - 399  
SMC3\_HUMAN - Q A T - - - - - Q E R T D L Y A - - - - - K Q G R G S Q F T S K E E R - D K W - - - - - I K K - - - - - E L K - 400  
SMC4\_HUMAN - K E K E S R E K E L M G F S K S V N E - - - - - A R S K M D V A K S E L D I Y - - - - - L S R H N - - - - - 514  
SMC5\_HUMAN - K E K K S V D D H I V R F D N L M N Q K - - - - - E D K L R Q R F R D T Y - - - - - 464  
SMC6\_HUMAN - - - - - E E L K K S T - - - - - D Q S L E P E R L E R Q - - - - - 407  
RAD50\_HUMAN - - - - - Q K Q I D E I R D K K T G L G R I E L K S E I L S K K Q N E L K N V K Y E L Q Q L E G S S D R I L - 475

SMC1A\_HUMAN - T L A Q E L E K F N R D Q K A D Q D R - - - - - L D L E E R K K V E T E A K I K Q K L R E I E N Q K R I - 435  
SMC1B\_HUMAN - T M T Q Q L E K L Q W E Q K T D E E R - - - - - L A F E K R R H G E V Q G N L K Q I E Q I E D H K K R I - 435  
SMC2\_HUMAN - - - - - G Q M M A C K N D - - - - - I S K A Q T E A K Q A Q M K L K H A Q Q E L K N K Q A E V - 437  
SMC3\_HUMAN - S L D Q A I N D K K R Q I A A I H K D - - - - - L E D T E A N K E K N L E Q Y N K L D Q D L N E V K A R V - 448  
SMC4\_HUMAN - T A V S Q L T K A K E A L I A A S E T - - - - - L K E R K A A I R D I E G K L P Q T E Q E L K E K E L - 562  
SMC5\_HUMAN - - - - - D A V L W L R N N - - - - - R D K F K Q R V - - - - - C E - - - - - P I M L T I N M K D N K N - 496  
SMC6\_HUMAN - - - - - K K I S W L K E R - - - - - V K A F Q N Q E - - - - - N S V N Q E I E Q F Q A I E K D K E E H - 445  
RAD50\_HUMAN - E L D Q E L I K A E R E L S K A E K N S N V E T L K M E V I S L Q N E K A D L D R T L R K L D Q E M E Q L N H H T T T R - 535

SMC1A\_HUMAN - E K L E E Y I T T S K - - - - - Q S L E E Q - K K L E G E L T D E E V E M - - - - - A K R R I - - - - - D E I N K E L N Q V - 480  
SMC1B\_HUMAN - E K L E E Y T K T C M - - - - - D C L K E K - K Q Q E E T L V D E I E K - - - - - T K S R M - - - - - S E F N E E L N L I - 480  
SMC2\_HUMAN - K K M D S G Y R K D Q - - - - - E A L E A V - K R L K E K L E A E M K K L N Y E E N K E E S L - L E K R R Q L S R D I G R L - 492  
SMC3\_HUMAN - E E L D R K Y Y E - - - - - V - K N K H D E L Q S E R N Y L W R E E N A E Q Q A L A A K R E D L E K K Q Q L L - 497  
SMC4\_HUMAN - Q K L T Q E E T N F K - - - - - S L V H D L F Q K V E E - - - - - A K S S L - - - - - 590  
SMC5\_HUMAN - A K Y I E N H I P S N - - - - - D - - - - - L R A F V F E - - - - - S Q E D M E V F L K E V - 527  
SMC6\_HUMAN - G K I K R E E L D V K - - - - - H A L S Y N - - - - - Q - - - - - R Q L K E L - 469  
RAD50\_HUMAN - T Q M E M L T K D K A D K D E Q I R K I K S R H S D E L T S L L G Y F P N K K - Q L E D W L H S K - - - - - S K E I N Q T - 590

SMC1A\_HUMAN - M E Q L G D A R I D R Q E S S R Q Q - - - - - R K A E I M E S I K - - - - - R L Y P G S V Y G R L I D L C Q P T Q K K Y Q I A V - 534  
SMC1B\_HUMAN - R S E L Q N A G I D T H E G K R Q Q - - - - - K R A E V L E H L K - - - - - R L Y P D S V F G R L F D L C H P I H K K Y Q L A V - 534  
SMC2\_HUMAN - K E - T Y E A L L A R F P N L R F A Y K D - P E K N W N - - - - - R - - - - - N C V K G L V A S L I S V K D T S A T T A L - 541  
SMC3\_HUMAN - R A A T G K A I L N G I D S I N K V L D H F R R K G I N Q H V Q - - - - - N G Y H G I V M N N F E C - E P A F Y T C V - 550  
SMC4\_HUMAN - - - - - A M N R S - - - - - R G K - - - - - V L D A I I Q E K K - - - - - S G R I P G I Y G R L G D L G - A I D E K Y D V A I - 633  
SMC5\_HUMAN - R D N - K K L R V N A V I A P K S S Y A D K A P S R S L N E L K Q Y G F F S Y L R - - - - - E - - - - - L - 569  
SMC6\_HUMAN - K D S - K T D R L K R F - G P N V P - - - - - A L L E A I D D A Y R Q G H F T Y K P V G P L G A C I H L R D P E L A L A I - 523  
RAD50\_HUMAN - R D R L A - - - - - K L N K E L A S S E Q - - - - - N K N H I N N E L K R K E E Q L S S Y E D K L F D V C G S Q - - - - - 636

SMC1A\_HUMAN - T K V L G K N M D A I I - - - - - V D S E K T G R D C - I Q Y I K E Q R G E P E T F L P L D Y L E V K P T D E - 583  
SMC1B\_HUMAN - T K V F G R F I T A I V - - - - - V A S E K V A K D C - I R F L K E E R A E P E T F L A L D Y L D I K P I N E - 583  
SMC2\_HUMAN - E L V A G E R L Y N V V - - - - - V D T E V T G K K L - L E - - - - - R G E L K R R Y T I I P L N K I S A R C I A P E - 589  
SMC3\_HUMAN - E V T A G N R L F Y H I - - - - - V D S D E V S T K I L M E F N K M N L P G E V T F L P L N K L D V R D T A Y P - 601  
SMC4\_HUMAN - S S C - C H A L D Y I V - - - - - V D S I D I A Q E C - V N F L K R Q N I G V A T F I G L D K M A V W A K M T - 682  
SMC5\_HUMAN - F D A P D P V M S Y L C C Q Y H I H E V P V G T E K T R E R - - - - - I E - - - - - 601  
SMC6\_HUMAN - E S C L K G L L Q A Y C C H N H A D E R V L Q A L M K R F Y - - - - - L P - - - - - 555  
RAD50\_HUMAN - - - - - D F E S D L D R L K - - - - - E E I E K S S K Q R - - - - - A - - - - - 657

SMC1A\_HUMAN - K L R - - - - - E L K G A K L V I D V I R Y E P - - - - - P H I K K A L Q Y A C G - - - - - N A L V C D N V E D A R - 626  
SMC1B\_HUMAN - R L R - - - - - E L K G C K M V I D V I K T Q F - - - - - P Q L K K V I Q F V C G - - - - - N G L V C E T M E E A R - 626  
SMC2\_HUMAN - T L R V A Q N L V G P D N V H V A L S L V E Y K - - - - - P E L Q K A M E F V F G - - - - - T T F V C D N M D N A K - 637  
SMC3\_HUMAN - - - - - E T N D A I P M I S K L R Y N - - - - - P R F D K A F K H V F G - - - - - D T L I C R S M E V S T - 640  
SMC4\_HUMAN - E I Q - - - - - T P E N T P R L F D L V K V K D - - - - - E K I R Q A F Y F A L R - - - - - K T L V A D N L D Q A T - 725  
SMC5\_HUMAN - - - - - R V I Q E - - - - - 606  
SMC6\_HUMAN - - - - - G T S R P P I I V S E F R N E I Y D V R H R A A Y H P D F P T V L T A L E I D N A V V A N S L I D M R - 606  
RAD50\_HUMAN - - - - - M L A G A T A V Y S Q F I T Q - - - - - 672

SMC1A\_HUMAN - R I A - - - - - F G G H Q R H K T V A L D G T L F Q K S G V - - - - - I S G G A S D L K A K A R R W - 666  
SMC1B\_HUMAN - H I A - - - - - L S G P E R Q K T V A L D G T L F L K S G V - - - - - I S G G S S D L K Y K A R C W - 666  
SMC2\_HUMAN - K V A - - - - - F D K R I M T R T V T L G G D V F D P H G T - - - - - L S G G A R S Q A A S I L T K F - 678  
SMC3\_HUMAN - Q L A - - - - - R A F T M D C I T L E G D Q V S H R G A - - - - - L T G G Y Y D T R K S R L E L Q - 679  
SMC4\_HUMAN - R V A - - - - - Y Q K D R R W R V V T L Q G Q I I E Q S G T - - - - - M T G G G S K V M K G R M G S S - 766  
SMC5\_HUMAN - - - - - T R L K Q I Y T A E E K Y V V K T S F Y S N K V I S S N T S L K V A Q F - 642  
SMC6\_HUMAN - G I E T V L L I K N N S V A R A V M Q S Q K P P K N C R E A F T A D G D Q V F A G R Y S S E N T - - - - - R P K F - 659  
RAD50\_HUMAN - - - - - L T D E N Q S C C P V C Q R V F Q T E A E - - - - - L Q E V I S D L Q S K L R L A P - 709

SMC1A\_HUMAN - - - - - D E K A V D K L K E K K E R L T E E - - - - - L K E - 687  
SMC1B\_HUMAN - - - - - D E K E L K N L R D R R S Q K I Q E - - - - - L K G - 687  
SMC2\_HUMAN - Q E L - - - - - K D V Q D E L R I K E N E L R A L E E E L A G L K N T A E K Y R Q L K Q - 717  
SMC3\_HUMAN - K D V - - - - - E E L G E L E A K L N E N - - - - - L R R - 702  
SMC4\_HUMAN - L V I - - - - - E I S E E E V N K M E - - - - - S Q L Q N D - - - - - S K K - 789  
SMC5\_HUMAN - L T V - - - - - T V D L E - Q R R H L E E Q L K E I H R K L Q A V D S G L I A L R E T S K - 681  
SMC6\_HUMAN - L S R - - - - - V D D S E - - - - - I S D L E N E V E N K T A Q I L N L Q - - - - - Q - 687  
RAD50\_HUMAN - D K L K S T E S E L K K K E K R R D E M L G L V P M R Q S I I D L K E K E I P E L R N K L Q N V N R D - - - - - I Q R L K N - 766

SMC1A\_HUMAN - QMKAKRKEAELRQVQSQAHG-LQMRLKYSQS----DLEQTKTRHLAL---NLQEK S-----735  
 SMC1B\_HUMAN - LMKTLRKETDLKQIQTLIQG-TQTRLKYSQN----ELEMIKKKHLVA---FYQE Q S-----735  
 SMC2\_HUMAN - QWEMKTEEADLLQTKL----QSSSYHKKQE---ELDALKKTIEESEETLKNTKEIQRK-768  
 SMC3\_HUMAN - NIERINNEIDQLMNQMQUIETQQRKFKASRDSILSEM KMLKEKRQQSEKTFMPKQ-----757  
 SMC4\_HUMAN - AMQIQEQK-----VQLEE-RVVKL R HSER---EMRNTLEKFTASIQRLIEQEEYLN-836  
 SMC5\_HUMAN - HLEHKDNE--LRQ-----K KELLER-700  
 SMC6\_HUMAN - HLSALEKD--IKH-----N-ELLKR-705  
 RAD50\_HUMAN - D I E E-----QETLLGTIMPEEESA KV-787

SMC1A\_HUMAN -----KLES E LANFGPRI---NDIKRIIQSREREMKDLKEKMNQVEDEVFEEFCREIGV-786  
 SMC1B\_HUMAN -----QLQSE L L NIESQC---IMLSEGIKERQRRRIKEFQEKIDKVEDDIFQHFCEEIGV-786  
 SMC2\_HUMAN - AEEKYEVLEN K M K N A E-----AERERELKDA-----QKKLDCAKTKA-805  
 SMC3\_HUMAN ---RSLQSL E A S L H A M E-----STRESLKAE-----LGT-783  
 SMC4\_HUMAN ---VQVKELEA N V L A T A P D K K K Q K L L E E N V S A F K T E Y D A V A E K A G K V E A E V K R L H-----888  
 SMC5\_HUMAN - KTKKRQLEQ K I S-----SKLGSLKLM-----721  
 SMC6\_HUMAN - CQLHYKELKM K I R-----KNISEIREL-----727  
 RAD50\_HUMAN - CL-----T D V T-----IMERFQMELKDVERKIAQQA AKLQGIDL D-----822

SMC1A\_HUMAN - RNIREFEE E K V K R Q N E I A K K R L E F E-----N Q K T R L-----817  
 SMC1B\_HUMAN - ENIREFEN K H V K R Q Q E I D Q K R Y F Y K-----K M L T R L-----817  
 SMC2\_HUMAN - DASSKKMKE--KQQEVEAITLELE-----ELKREHT--SY--836  
 SMC3\_HUMAN - DLLSQLSL E --DQKRVDALNDEIR-----QLQENR--QLLN-816  
 SMC4\_HUMAN - NTIVEINN K L K A Q Q D-----K L D K I-----909  
 SMC5\_HUMAN -----E-----QDTCNLEEE-----731  
 SMC6\_HUMAN -----E N I E E H Q S V D I A T L E D E-----744  
 RAD50\_HUMAN - RTVQQVNQ E K Q E K Q H K L D T V S S K I E L N R K L I Q D Q Q E Q I Q H L K S T T N E L K S E K L Q I S T N L Q-----882

SMC1A\_HUMAN --GIQLDFEKNQLKEDQDKVHMWEQT V K K D E N E I E K-----851  
 SMC1B\_HUMAN --NVQLEYSRSHLKKLKNKINTLKET I Q K G S E D I D H-----851  
 SMC2\_HUMAN --KQQL EAVNEAIKSYESQIEVMAAE V A K N K E S V N K A Q E E V T K-----QKEVITAQD TVI-889  
 SMC3\_HUMAN - ERKLEGIITRVETYLNE-----N L R K R L D Q V E---QELNE-----LRE--TEGGTVL-859  
 SMC4\_HUMAN ---NKQLDECASAITKAQVAIKTADRNL Q K A Q D S V L R-----943  
 SMC5\_HUMAN -----ER--KASTKIKEIN--V Q K A K L V T E L T N L I K I C T S L H I Q K V D L I L Q-----773  
 SMC6\_HUMAN -----AQ--ENKSKMMV E E H M E Q Q K E N M E H L K S L K I E A E N K Y D A I K F K I N-----788  
 RAD50\_HUMAN - RRQQL E E Q T V E L S---TEVQSLYREIK D A K E Q V S P L E T T L E K---FQKEKEELINKKNTSN-937

SMC1A\_HUMAN ---LKKEEQRHMKI I D E T M A Q L Q D L K N Q H L A K K S E V N D K N H E M E E I R K K-----LGGAN-902  
 SMC1B\_HUMAN ---LKKAEENCLQTV N E L M A K Q Q L K D I R V T Q N S S A E K V Q T Q I E E E R K K-----FLAVD-902  
 SMC2\_HUMAN - KAKYAEVAKHKEQN D S Q L K I K E L D H N I S K H K R E A E D G---AAKVS K M-----934  
 SMC3\_HUMAN - TATTSELEAINKRV K D T M A R S E D L D N S I D K T---EAG---IKELQ K S-----900  
 SMC4\_HUMAN ---TEKEIKDTEKEV D D L T A E L K S L E D K A A E V V K N T N A A E E S L P E I Q K E H R N L L Q E L K V I Q-----1001  
 SMC5\_HUMAN ---NTTVISEKNKLE S D Y M A A S S Q L---RLTEQHF-----IELDENRQRL L Q---KCK-817  
 SMC6\_HUMAN ---QLSELA--DPLK D E L N L A D S V D N Q K R G R H Y E E K Q K E H L D T L N K K K R E L D M K E K E L E-844  
 RAD50\_HUMAN - KIAQDKLNDIKEKV K N I H G Y M K D I E N Y I Q D G K D D Y K K Q---KETELN K V---IAQLSECE-991

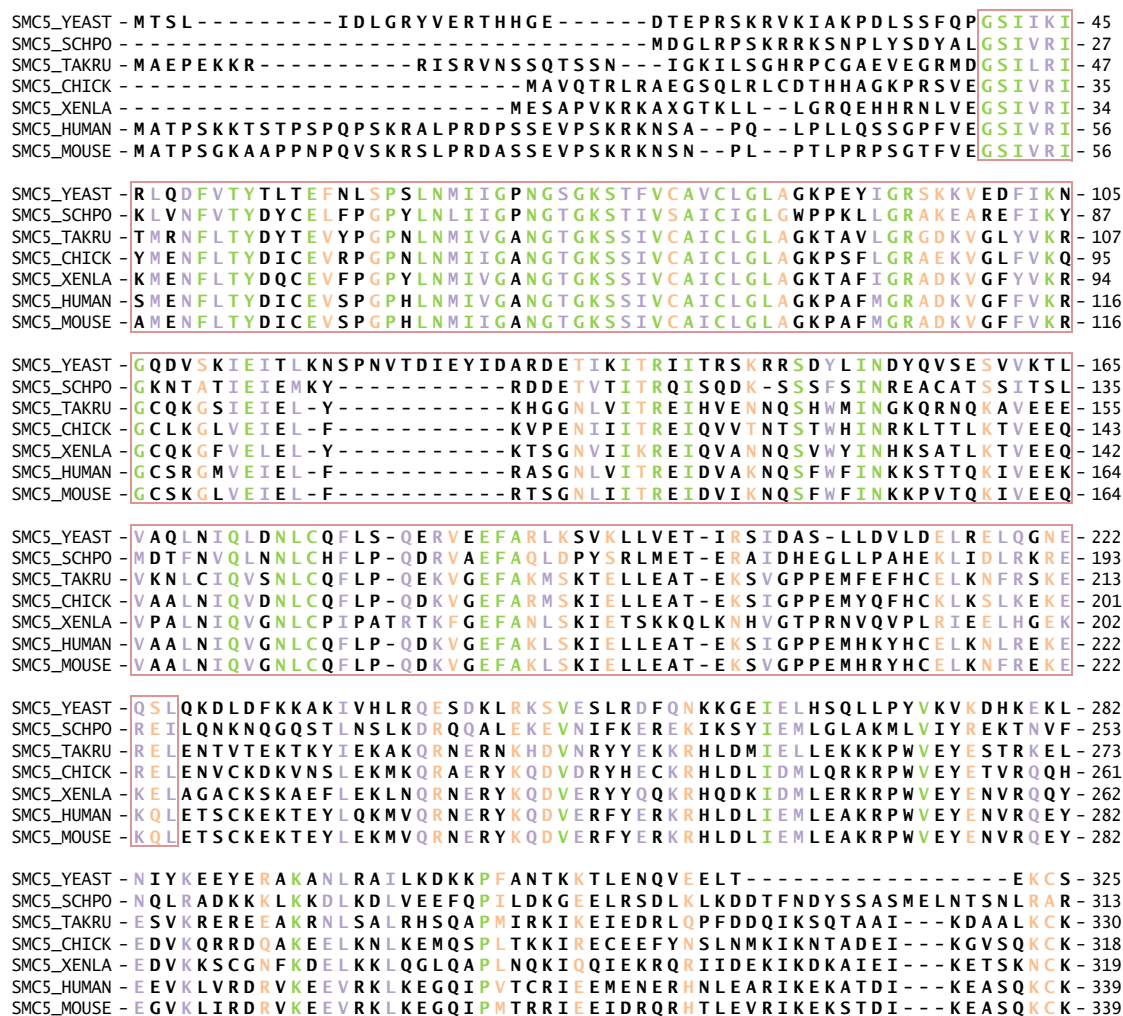
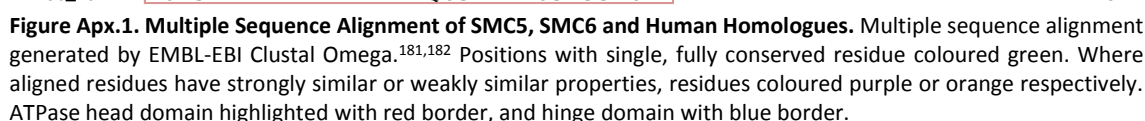
SMC1A\_HUMAN - KEMTHLQKEVTAIETKLEQKRSRDRHNL L Q A C K M Q D I K L P L S K G T M D D I S Q E E G S S Q G E D S-962  
 SMC1B\_HUMAN - REVGKLQKEVVS I Q T S L E Q K R L E K H N L L L D C K V Q D I E I I L L S G S L D D I E V E M G T E A E---960  
 SMC2\_HUMAN -----L-----KDYDWIN-AER-----945  
 SMC3\_HUMAN -----M-----ERWKN-MEKEHMDAINHDTKELE--K-924  
 SMC4\_HUMAN - ENEHALQKDALS I K L K L E Q I D G---HIAEHNSKI KY W H K E I S K I S L H P I E-----1048  
 SMC5\_HUMAN - E L M K R---AR-----824  
 SMC6\_HUMAN - E K M S Q---AR-----851  
 RAD50\_HUMAN - K H K E K I N E D M R L M-----R Q D I D T Q K I Q E R W L Q D N L---T L-----1024

SMC1A\_HUMAN - VSGSQRISS I Y A R E A L I E I -D Y G D L C E D L K D A---Q A E E E I K Q E---MNTLQQKLNEQQS-1015  
 SMC1B\_HUMAN ---STQATIDIYEKEEAFEI-DYSSLKEDLKAL--QSDQEIEAH---L R L L L Q Q V A S Q E D-1011  
 SMC2\_HUMAN -----H L F G Q P N S A Y D F K T N N P K E A G Q R---L---Q K L Q E M K E-977  
 SMC3\_HUMAN - M T N R Q G M L L K K K E E C M K K I R E L G S L P Q E A F E K Y---Q T L S L K Q L---F---R K L E Q C N T-974  
 SMC4\_HUMAN -----D N P I E E I -S-VLSPEDLEAI--KNPDSITNQ---I A L L E A R C-----1083  
 SMC5\_HUMAN -----Q V C N L G A E---Q T L P Q E---Y Q T Q V P T I P N G H N S S L P M V F Q D L P N T L D-866  
 SMC6\_HUMAN -----Q I C P E R I E -V E K S A-----S I L D K-----869  
 RAD50\_HUMAN -----R K R N E E L K E V E E-----E R K Q H L K E M G Q M Q V L Q M K S E H Q K L E E N I D-1065

SMC1A\_HUMAN - V L Q---R I A A P N M K A-----M E K L E S V R D K F Q E T S D E F E A A R K R A K K A K-1056  
 SMC1B\_HUMAN - I L L---K T A A P N L R A-----L E N L K T V R D K F Q E S T D A F E A S R K E A R L C R-1052  
 SMC2\_HUMAN - K L G---R N V N M R A-----M N V L T E A E E R Y N D L M K K K R I V E N D K S K I L-1016  
 SMC3\_HUMAN - E L K---K Y S H V N K K A-----L D Q F V N F S E Q K E K L I K R Q E E L D R G Y K S I M-1015  
 SMC4\_HUMAN -----H E M K P N L G A-----I A E Y K K K E E L Y L Q R V A E L D K I T Y E R D S F R-1121  
 SMC5\_HUMAN - E I D A L L T E E R S R A S C F T G L-----N P T I V Q E Y T K R E E E I E Q L T E E L K G K K V E L D Q Y R-918  
 SMC6\_HUMAN - E I N R L -R Q K I Q A E H A S H G D-----R E E I M R Q Y Q E A R E T Y L D L D S---K V R T L K K F I-916  
 RAD50\_HUMAN - N I K---R N H N L A L G R Q K G Y E E E I I H F K K E L R E P Q F R D A E E K Y R E M M I V M R T T E L V N K D L D-1122

SMC1A\_HUMAN - Q A F E Q I K K E R F D R F N A---C F E S V A T N I D E I Y K A L S R N S S A Q A F L G P E N P E-----1104  
 SMC1B\_HUMAN - Q E F E Q V K K R R Y D L F T Q---C F E H V S I S I D Q I Y K K L C R N N S A Q A F L S P E N P E-----1100  
 SMC2\_HUMAN - T T I E D L D Q K K N Q A L N I---A W Q K V N K D F G S I F S T L L P G A N A M L A P P E G Q T V---1064  
 SMC3\_HUMAN - E L M N V L E L R K Y E A I Q L---T F K Y V S K N F S E V F Q K L V P G G K A T L V M K K G D V E G S Q S Q D E G-1071  
 SMC4\_HUMAN - Q A Y E D L R K Q R L E F M A---G F Y I I T N K L K E N Y Q M L T L G G D A---E L E L V D S L-----1167  
 SMC5\_HUMAN - E N I S Q V K E R W L N P L K E---L V E K I N E K F S N F F S S M Q C A G E V D L H T E N E E-----964  
 SMC6\_HUMAN - K L L G E I M E H R F K T Y Q Q F R R C L T L R C K L Y F D N L L S Q R A Y C G K M N F D H K N E T-----966  
 RAD50\_HUMAN - I Y Y K T L D Q A I M K F H S M---K M E E I N K I I R D L W R S T Y R G Q D I E Y I E I R S D-----1168







SMC5\_YEAST - LKTDEF LKAKKII-NEIFEKLNTRI DEVIKKNQNEYYRGRTKKLQATIISTKEDFLRSQ-384  
 SMC5\_SCHPO - ASFSNFMENEKKLYEKVNTNRTLLRN-----A-NLTLENAQQSV-351  
 SMC5\_TAKRU - QK-QDQLDRKQKEIEDINQAFKLKMEEDDHQ-----KRISNT-RRIEDLRTLE-378  
 SMC5\_CHICK - EK-QDALEMKDKQISEINQALRMKKDEEVDK------KKILSA-YKMIIDWNNEL-366  
 SMC5\_XENLA - QK-QDELEQKDKQIEEVQQSRLRVKDAEQRQ------KKIGNI-RKMIIDWEKEL-367  
 SMC5\_HUMAN - QK-QDVIERKDKHIEELQQALIVKQNEELDRO------RRIGNI-RKMIIDLQNEL-387  
 SMC5\_MOUSE - QR-QDLIERKDRQIKELQQALTVKQNEELDRO------KRISNT-RKMIIDLQSEL-387

SMC5\_YEAST - EILAQTH--LPEKSVFEDIIDIKRKEINKEGEIRDLS EIDAKANAINHEMRSIQRQAES-442  
 SMC5\_SCHPO - KSLTERQGPRPSDNGVQDLQEKMEVNAE-----KLQHENEKLESSEHLSGSI RTLKAQ-404  
 SMC5\_TAKRU - AKVE-----DQPDVTPRINDVNSE-----LRRNQIERARIDGKCELCREKDN-421  
 SMC5\_CHICK - NTVT-----DCENLQPDIDAVNNE-----LKHVQEEERANIDSDIGDVTTKEIN-409  
 SMC5\_XENLA - SAMT-----NQENIQPEMDSINTD-----FRRIQDDKANIESEMTDLRMEKDN-410  
 SMC5\_HUMAN - KTT-----NCENLQPDIDAVNNE-----LRRIQDEKALCEGEIIDKRRERET-430  
 SMC5\_MOUSE - KTA-----NCENLQPDIDAVNND-----LRRVQEEKALCEGEIIDKQREKEM-430

SMC5\_YEAST - KTKSL-TTTDKIGIL-----NQQDQLKEVRDAVLMVRE-HPPEMKDKILEPPIMT-489  
 SMC5\_SCHPO - KLIDLDNIKR ELSYYNDATKR KLD FMSAPGWEDAYQTYQLLKEYE SAFAEPAYGPIYMN-464  
 SMC5\_TAKRU - AFAQCRSLQK KLDNMNLMKV KEE--KLGRHRDTHAALQLWRQNNRFRGNVYEPMLLE-479  
 SMC5\_CHICK - QERENGRIIDRIGQLNNIIVK KEE--TLQARFRDTHSALMWLRKNKHKFKKEVCEPMLLT-467  
 SMC5\_XENLA - LERKKKEKANRIKQFDNLMNL KEE--KLKRMYSDTYNVVLKKNKDFKRVCEPMLLE-468  
 SMC5\_HUMAN - LEK EKKSVDDHIVRFDNLMNQ KED--KLQRFRD TYDAVLWLRRNNRDKFKQRVCEPIMLT-488  
 SMC5\_MOUSE - LEKQRRSVSDHITRFDNLMNQ KED--KLQRQYRDTYDAVLWLRRNNRDRFKQRVCEPIMLT-488

SMC5\_YEAST - VSAINAQFAAYLAQCVDYNTSKALT VVDSYKLFANPILDKFKV--NL-----RELSS-541  
 SMC5\_SCHPO - LKCKEKGFAALIEGFRDTFTFIMSNYDLKLMDLITSKTKYTP TIREFSSERKKKI-524  
 SMC5\_TAKRU - INVVDHRRFAKYVENHISFDLRAV FQRKEDMEIFMSEVRDKMNLKVN--SISAPQSR-536  
 SMC5\_CHICK - INVVDNKHAKYVENHISNDMKAFF FESQEDMELFLVELRDRQKL RVN--AVCAPDKSC-524  
 SMC5\_XENLA - INMKDQRHAKYVENHIPMNDKAFV FESQEDMQVFLKEVRDKQNL RVN--TVCAPQEPY-525  
 SMC5\_HUMAN - INMKDNKNKAYIENHIPSNDLRAFV FESQEDMEVFLKEVRDNKKL RVN--AVIAPKISSY-545  
 SMC5\_MOUSE - INMKDNKNKAYIENHISNDLRAFV FESQEDMEIFLREVRDNKKL RVN--AVIAPKISSY-545

SMC5\_YEAST - ADTTPPVPAETVRDLGFEGLSDFITGDKRVMMKLCQTSKIHTIPVSRRLETPAQIKKLI-601  
 SMC5\_SCHPO - EDFEPPCSREKLQSF GFDGYVIDFLEGPVEVLVALCHMLKIHQIPAKRELPPASVNALN-584  
 SMC5\_TAKRU - SKAQPSQNIEDLRRFGFFTYLRFMDADPEVMSYLCQQYNVHNVPVGTETKTMTIRQVIE-596  
 SMC5\_CHICK - AETLPSTPIEELHRYGFFSYLRELFADPLPVM SYLCSQYHVHEVPVGTETKTRNMIRVJK-584  
 SMC5\_XENLA - AEQRPKRPITDLKQYGFSSYLRELFADAPYPMNYLCYQYKVHEVPVGTETQTRSMIEKVIK-585  
 SMC5\_HUMAN - ADKAPSRSLNELKQYGFSSYLRELFADAPDPVMSYLCQQYHIEHVPVGTETKTRERIERVJK-605  
 SMC5\_MOUSE - ADKAPSRSLNDLKQYGFSSYLRELFADAPDPVMSYLCQQYHIEHVPVGTETKTRERIERVJK-605

SMC5\_YEAST - TPR-PNGKILFKRIIHGNRLVDIKQSA YGSKQVFPTDVS IKQT-NFYQGSIMSNEQKIRI-659  
 SMC5\_SCHPO - NFRLANGDPVLKTYLAGSSIHVFRSAYGDREITRR-TDPLPSRSIYFSENVEMDLVKRR-643  
 SMC5\_TAKRU - ELNLRV-----LFTLDERYMLKR SVYSKMISTIN-SPVNP S--QYLSIAVDAEKRQL-646  
 SMC5\_CHICK - ETKLKQ-----IYTAEEKYTIKVS TYTKLSFSTN-MCLRPA--QFLNYVDTDERQL-634  
 SMC5\_XENLA - ETDLRQ-----MYTAEEKYVTKK SVYSKLISSN-VSLKGA--QFLT VTVDAEQRQV-635  
 SMC5\_HUMAN - ETRLKQ-----IYTAEEKYVVKTSFYSNKVISSN-TSLKVA--QFLT VTVDAEQRRL-655  
 SMC5\_MOUSE - ETRLKQ-----IYTAEEKYVVKTSVYSNKVISSN-TSLKVA--QFLT VTVDAEQRRL-655

SMC5\_YEAST - EN EIIINLKNEYNDRKSTLDALSNQKSGYRHELSE LASKNDDINREAHQLNEIRKKYTMRK-719  
 SMC5\_SCHPO - EEQLNAQLS-----QLENLQNEERK-----LQEKVNEHESLLSRTNDLSTLRKE-689  
 SMC5\_TAKRU - EQELNACELRFREIDERLKA LQRETA V-----LDRD-----NELL-----AEK-685  
 SMC5\_CHICK - ENQQQNIKHILQSLDKQLMTLCERQKH-----LECRD-----NELR----QKQ-673  
 SMC5\_XENLA - VDQLKEIERKCS TIETSM EQLAERQRS-----LDRD-----NELR----LRK-674  
 SMC5\_HUMAN - EEQLKEIHRKLLQAVDSGLIALRETSKH-----LEHKD-----NELR----QKQ-694  
 SMC5\_MOUSE - EEQLKEMNRQLEAVDSGLAALRDTNRH-----LELKD-----NELR----LKK-694

SMC5\_YEAST - STIET-----LRE-----KLDQLKRE-----ARKDVSQKIKDIDDDQIQQLLLKQR-759  
 SMC5\_SCHPO - DEKLIPHEWQQLQERIEHQ TLLLRQREKVPEQFAAEI-EKNEDIRKEN-----737  
 SMC5\_TAKRU - KKLSELKGKKRQLEQKISTKQDSLRQMEQNVTDLKKIEEETKEKVS AVNSQKVT-----739  
 SMC5\_CHICK - KELLERSRRKQLESKIAVKYDSIRQLEQNPINLEKESQAKVKIRAINIQKAK-----727  
 SMC5\_XENLA - KTIEMKTKKRQLEQKISTKYDSLNRLEQDNLNLEEVQQANNRINKNLNVQKAK-----728  
 SMC5\_HUMAN - KELLERKTKKRQLEQKISSKLGLSKLMEQDTCNLEEEERKASTKIK EINVQKAK-----748  
 SMC5\_MOUSE - KELLERKTKKRQLEQKISSKLASIRLMEQDTCNLEEEERKASTKIK EINVQKAK-----748

SMC5\_YEAST - HLLSKMASSMKS LKNCQKELISTQILQFEAQNM DVSMNDVIGFFNER-----806  
 SMC5\_SCHPO - F-E---ALMNSVLKVKENSIKATNN-F-EKMLGSR LNVTEAKYKLEKHEMDANQVNARL-790  
 SMC5\_TAKRU - IVK---AFIASIKLKATLTMEKVYLSLEMMGLSAEKTLEHDFREGASLLRS--MDQRC-793  
 SMC5\_CHICK - LVT---ELMCHIKNYVSLNICKADLI LQSTAVDAEKNRLEAEYKAASVELRA--SEQRF-781  
 SMC5\_XENLA - LVK---DLLELMKECTSLSEKVELALQSTASISEKNKIESDYKSATSQ LRE--LKNQY-782  
 SMC5\_HUMAN - LVT---ELTNLIKICTSLHIQKVVDLI LQNTTVISEKNKLES DYMAASSQLRL--TEQHF-802  
 SMC5\_MOUSE - LVT---ELTGLVKICTSFQIQKVVDLI LQNTTVISEKNKLEADYMASSSQLRV--TEQQF-802

SMC5\_YEAST - EADLKSQYEDKKKFVKEMRDT--PEFQSWMR EIRSYDQDTKEKLNKVAEKYEEEGN-860  
 SMC5\_SCHPO - TEVQDR LK DITDKL---ASARE-DAMSLYGSVVD SLQTQSS-----D-828  
 SMC5\_TAKRU - SLEQRKVQLTEQGGQMKRAKSI CNMGPNDLSLELRNVRV-----YVIPPYLCVPS-846  
 SMC5\_CHICK - LELDERKRIILT ENCKELLKKARQMCNMNL DQHLPEFQT-----820  
 SMC5\_XENLA - DGI EAKKLHLL ENCKGLLRKARQACNLGPNQAVPQDFQT-----821  
 SMC5\_HUMAN - IELDENRQRL LKCKELMKRARQVCNLSAQETLPQ EYQTQVP-----T-IPNGHNS-853  
 SMC5\_MOUSE - IELDDNRQRL LKCKELMKRARQVCNLSADQAVPQ EYQTQVP-----T-IPNGHNS-853

SMC5\_YEAST - F--NLSFQDVLDKLESEIAMVNH-----DESAVTILDQVTAELRELEHTVPQQ-907  
 SMC5\_SCHPO - RQTAITELNEEFATSS EVDNKISIEE-TKLKFMNVNSYVMEQYDARKK EIEELESKMSDF-887  
 SMC5\_TAKRU - PLMAFAKLP--DTPDDIDSM LNEERSRSECTGLSEN VDEYNRS DQ EIKEL ENELEEK-903  
 SMC5\_CHICK - A FQTL P--DTLEEIDAF LNEERSRVCSFTGLSASVVEECSKQMEIQKLMESEEN-874  
 SMC5\_XENLA - AFQSLP--ESLDEIDAM LNEERSRASCFTGLTASVDDYNKRTEI QEVTEELNRK-875  
 SMC5\_HUMAN - LPMVFQDLP--NTLDEIDAL LTEERSRASCFTGLNP TIVQEYTKRE EIEEQLTEELKKG-910  
 SMC5\_MOUSE - PPMAFQDLP--NTLDEIDAL LTEERSRASCFTGLNP SVVEEYSKRE EIEEQLTEELQKG-910

SMC5_YEAST	-	SKDLETIKAKLKEDHAVLEPKLDDIVSKISARFARLFNNVGSAGAVRLE--KPKDYAEWK	-	965
SMC5_SCHPO	-	DQSVEELQDEMNSIKEDWVSKLEENVQCISDRFSKGMGMYAGEVRLG--KSDDYDKWY	-	945
SMC5_TAKRU	-	KNALESYRQNISEAERWLNPLKQLVEQINEKFTAFFRSMNCAGEVDLHSEKEEDYDKYG	-	963
SMC5_CHICK	-	KKELDDYKQSSISKIKERWLNPLKKMIESINEKFSGFFSSMESVGEVDLHVENEEYDKYG	-	934
SMC5_XENLA	-	KLELDYRKNISQVKEWLNPLKQLIEKINDQSSFFSSMQCVGEVDLHTEKEEYDKYG	-	935
SMC5_HUMAN	-	KVELDQYRENISQVKEWLNPLKQLVEKINEKFSNFFSSMQCAGEVDLHTEKEEYDKYG	-	970
SMC5_MOUSE	-	KVELDEYRENISQVKEWLNPLKQLVEKINEKFSNFFSSMQCAGEVDLHTEKEEYDKYG	-	970
SMC5_YEAST	-	IEIMVKFRDNAPLKKLDSHTQSGGERAVSTVLYMIALQEFSTAPFRVVDENQGMDSRNE	-	1025
SMC5_SCHPO	-	IDILVQFREEEGLQKLTGQRQSGGERSVSTIMYLLSLQGLAIAPFRVVDENQGMDSRNE	-	1005
SMC5_TAKRU	-	IRIRVKFHSNTQLHELTFFHQSGGERSVSTMLYLMALQELNRCPPFRVVDENQGMDSRNE	-	1023
SMC5_CHICK	-	IRIRVKFHNFTDLHELTTPYHQSGGERSVSTVLYLMALQELNRCPPFRVVDENQGMDSRNE	-	994
SMC5_XENLA	-	IRIRVKFRSSTQLHELTTPHHQSGGERSVSTMLYLMALQELNRCPPFRVVDENQGMDSRNE	-	995
SMC5_HUMAN	-	IRIRVKFRSSTQLHELTTPHHQSGGERSVSTMLYLMALQELNRCPPFRVVDENQGMDSRNE	-	1030
SMC5_MOUSE	-	IRIRVKFRSSTQLHELTTPHHQSGGERSVSTMLYLMALQELNRCPPFRVVDENQGMDSRNE	-	1030
SMC5_YEAST	-	RIVHKAMVENACAENTSQYFLITPKLLTGLHYHEKMRTHCVMAGSWIPNPSEDPKMIHFG	-	1085
SMC5_SCHPO	-	RVVHRHIVNSVCDNAVSQYFLVTPKLLPDLYHRNLKVLICNGAWLPATFRTSLSTYFE	-	1065
SMC5_TAKRU	-	RRVFDIVVGTAACKERTSQYFFITPKLLQNLKYAEEMTVLCVHNGAYMLPPNQWDDKALR	-	1083
SMC5_CHICK	-	RRVFEMFVKTAACKERTSQYFLITPKLLQNLTYNEKMTLLFVYNGPFMLEANKWNLKSCFR	-	1054
SMC5_XENLA	-	RRVFEMVVKTAACKERTSQYFFITPKLLQNLTYAEKMTVLFVYNGPFMLEPTKWNLKAHR	-	1055
SMC5_HUMAN	-	RRVFEMVVNTACKERTSQYFFITPKLLQNLTPYSKMTVLFVYNGPFMLEPNTWNLKAFQR	-	1090
SMC5_MOUSE	-	RRVFEMVVNTACKERTSQYFFITPKLLQNLTPYSKMTVLFVYNGPFMLEPNRWNLKAQR	-	1090
SMC5_YEAST	-	ETSNYSD---	-	1093
SMC5_SCHPO	-	KLKKSALISS	-	1076
SMC5_TAKRU	-	RCLQRKAKA--	-	1092
SMC5_CHICK	-	RRRRLGRMDEQ	-	1065
SMC5_XENLA	-	RRRRAVADQ-	-	1065
SMC5_HUMAN	-	RRRRIITFTQPS	-	1101
SMC5_MOUSE	-	RRRRIITFTQPQ	-	1101

**Figure Apx.2. Multiple Sequence Alignment of SMC5 Orthologues.** Multiple sequence alignment generated by EMBL-EBI Clustal Omega.<sup>181,182</sup> Positions with single, fully conserved residue coloured green. Where aligned residues have strongly similar or weakly similar properties, residues coloured purple or orange respectively. ATPase head domain highlighted with red border, and hinge domain with blue border.

SMC6_YEAST	-	MI-----S--TTISGRPI-EQVDD---ELLSL-----TAQQENEEQQQQR	-	35
SMC6_SCHPO	-	MTTELTNVSLEEAITTEKSEN--RRKRDSDVLQT--EEVDLSNVKRRIRASRNQDNRPQR	-	58
SMC6_TAKRU	-	MS-----KRKSIETHDKSTKAR-----PV	-	20
SMC6_XENLA	-	MG-----KRKDGSPAPISSQRRKKQREQVEDDYPDEEDYG---QAGPSVSN DYGQRKKQR	-	50
SMC6_HUMAN	-	MA-----KRKEENFSSPKNAKRPRQEEL-----	-	23
SMC6_MOUSE	-	MA-----KRKEENFCSPENAKRPRQEEL-----	-	23
SMC6_YEAST	-	KRRRHQFAPMTQFNSNTLDEDSGFRSSSDVATADQDNFLEESPSGYIKKVIILNFMCH	-	95
SMC6_SCHPO	-	LQRSS--SLIEQVRGN--EDGEN--DVLNQTRRETNSNFDNRVGVICTHLVNFMCHDS	-	110
SMC6_TAKRU	-	E-----EEEDSEDA--DDRQDQVERLS--PQVLPSSGDVSDVGIVKSTILNFMCHAN	-	68
SMC6_XENLA	-	K-----EADNSHTD--SDYENSVCASSQRVSASQSGTGDDVGIIESIFLRNFMCHSM	-	100
SMC6_HUMAN	-	-----EDFDKD-----GDEDECKG-----TTLTAAEVGIIESITHLKNFMCHSM	-	61
SMC6_MOUSE	-	-----EDFDKD-----GDEDECTISFTNGTSTLTAAEVGIIESITQLNFMCHSM	-	67
SMC6_YEAST	-	FE-LELGSRLNFIVGNNGSGKSAILTAITIGLGAKASETNRGS SLKDLIREGCYSAKITIL	-	154
SMC6_SCHPO	-	LK-INGFPRINFVIGHNGSGKSAILTLTICLGAKASNTNRAPNMKSLVKQKGYARISV	-	169
SMC6_TAKRU	-	LGPFAFGSNVNFIVGNNGSGKSAILTGLIVALGGNAQATNRGS SLKGFVKEGESFAVSI	-	128
SMC6_XENLA	-	LGPFRFGPNVNFVIGHNGSGKSAVLTALIVGLGGKAAITNRGS SLKGFVKEGQTFAEISI	-	160
SMC6_HUMAN	-	LGPFKFGSNVNFVGHNGSGKSAVLTALIVGLGGRVATNRGS SLKGFVKDQNSADISI	-	121
SMC6_MOUSE	-	LGPFKFGSNVNFVGHNGSGKSAVLTALIVGLGGKAVATNRGS SLKGFVKAGQNSADISI	-	127
SMC6_YEAST	-	HLDNSKYGAYQQGIFGNEIIVERIIRKDPASFSLRSENGKEISNKKKDIQTVDYFVSVP	-	214
SMC6_SCHPO	-	TISNRGFEAYQPEIYGKSIITERTIRREGSEYRLRSFNGTVISTKRDELDNICDHMGLQ	-	229
SMC6_TAKRU	-	TLNNIGKDAYKPEVYQQAIVIDQKITREGIRTYKLKSQSGHIISTKKEDLVTILDYNNIQ	-	188
SMC6_XENLA	-	TLRNRGQDAYKPDVFGNSITVQQLRTTDCSRITYKLKSAFGAVVSNKKEELTAILDHFN	-	220
SMC6_HUMAN	-	TLRNRGDDAFKASVYGNISILIQHISIDGSRSYKLKSATGSVVSTRKEELIAILDHFN	-	181
SMC6_MOUSE	-	TLRNRGDDAFRANVYGDSIVVQHHISVDGSRSYKLKSEKGTVVSTRKEELIAILDHFN	-	187
SMC6_YEAST	-	VSNPMCFLSQDAARSFLTASTSQDKYSHFMKGTLTLEITENLLYASAIHDSAQENMALHL	-	274
SMC6_SCHPO	-	IDNPMNILTQDTARQFLGNSSPKEKYQLFMKGIIQLKQLEENYSLIEQSLINTKNVLGNKK	-	289
SMC6_TAKRU	-	VNNPVTILTQEMSKYFLHSKGGAEKYKFFMKATQLEQMKDDFVHIKSTKSVTDQVEQS	-	248
SMC6_XENLA	-	VNDPVS VLTQEMSKHFLQSKNESDKYKFFMKATQLEQMKEDYSYIMETKSRTHDQVENG	-	280
SMC6_HUMAN	-	VNDPVS VLTQEMSKQFLQSKNEGDKYKFFMKATQLEQMKEDYSYIMETKERTKEQIHQGE	-	241
SMC6_MOUSE	-	VNDPVS VLTQEMSKQFLQSKNEGDKYKFFMKATQLEQMKEDYSYIMETKERTKEQINQGE	-	247
SMC6_YEAST	-	ENLKS LKA EYEDAKKLLRELN---QTSDLNERKMLLQAKSLWIDVAHNTDACKNLEN	-	329
SMC6_SCHPO	-	TGVSYLA---KKEEYKLLWEQSRETNLHLLLEQKKGEMVWAQVVEVEKELLAEKEF	-	345
SMC6_TAKRU	-	ECLKDLKRDLKEEDRYKSLA---SVNEMTKLEELKKQMAWALVGEVEKEFEPMKEL	-	304
SMC6_XENLA	-	ERLRDLRQECIQKEERFKSIA---SLGEMKEKLEDLKNKMAWALVTESEKQIKPFIEQI	-	336
SMC6_HUMAN	-	ERLTE LKRQCVEKEERFQSI A---GLSTMKTNLES LKHEMAWAVVNEIEKQLNAIRDNI	-	297
SMC6_MOUSE	-	ERLTE LKRQCLEKEERFQNI A---GLSTMKTNLEY LKHEMAWAVVNEIEKQLNAIRDNI	-	303

SMC6\_YEAST - - - - - I S G I Q Q K V D E V T E K I R N R Q E K I E R Y T S D G T T I E A Q I D A K V I Y V N E K D - 376  
SMC6\_SCHPO - Q H A E V K L S E A K E N L E S I V T N Q S D I D G K I S S K E E V I G R A K G E T D T T K S K F E D I V K T F D G Y R - 405  
SMC6\_TAKRU - E S D R C A T N K F N E K V D E W K K K V E V A E G K Q K Q S - - - - - H E Q L E I T Q Q V S E L Q - 350  
SMC6\_XENLA - S T E E G R T V K Y E Q K I E E C Q G K V I N A E E K F R A K - - - - - Q E E L D K I T Q E A V A L K - 382  
SMC6\_HUMAN - K I G E D R A A R L D R K M E E Q Q V R L N E A E K Q Y K D I - - - - - Q D K L E K I S E E T N A R A - 343  
SMC6\_MOUSE - K I G E E R A A K L D R K M E E Q Q V R L N D A E K K Y K D I - - - - - Q D K L E K I S E E T N A R A - 349

SMC6\_YEAST - S E H Q N A R E L L R D V K S R F E K E K S N Q A E A - - Q S N I D Q G R K - - - K V D A L N K T I A H L E E E L T K E - 431  
SMC6\_SCHPO - S E M N D V D I Q K R D I Q N S I N A A K S C L D V Y R E Q L N T E R A R E N N L G G S Q I E K R - - - - - - 454  
SMC6\_TAKRU - S K C T - - - E F K T E V Q R R N A D L K S C E V T V H R H K A I L R D L E K - - - D K A Q I S S K I N D L S L S I S Q A - 405  
SMC6\_XENLA - P Q G I - - - G L K E D V Q K K R K S Y N E S E V L Y N R H R M E L K R L E R - - D A E Q L H K R I E E L K K S A D N D - 437  
SMC6\_HUMAN - P E C M - - - A L K A D V V A K K R A Y N E A E V L Y N R S L N E Y K A L K K - - D D E Q L C K R I E E L K K S T D Q S - 398  
SMC6\_MOUSE - P E C M - - - A L K T D V I A R T R A F N D A E V L Y N R S L N E Y K A L K K - - D G E Q L C K R I E E L K K S T D Q S - 404

SMC6\_YEAST - M G G D K D Q M R Q E L E Q L E K A N E K L R E V N N S L V V S L Q D V K N - - - E E R D I Q H E R E S E L R T I S K S - 488  
SMC6\_SCHPO - - A N E S N N L Q R E I A D L S E Q I V E L E S K R N D L H S A L L E M G G - - - - - N L T S L L T K D S I A N K - 506  
SMC6\_TAKRU - T G A E S Q A R M E R I A Q I E A A L E D L T H T T S T L G Q Q I E Q Y Q H S Y R H A I E E G Q G K M K R E L E G L Q K S - 465  
SMC6\_XENLA - S E S E K M A R Q K E I N Q I R E R M K A L H D K D I T N Q Q I H Q F Q Q A I E K Y E E R A R I G N E R N I K Q R - 497  
SMC6\_HUMAN - L E P E R L E R Q K K I S W L K E R V K A F Q N Q E N S V N Q E I E Q F Q Q A I E K D K E E H G K I K R E E L D V K H A - 458  
SMC6\_MOUSE - L E P E R L E R Q K R I C W L K E K V K A L Q D Q E H T V N Q E A E Q F E Q A I E K D K Q E H G R V R K E D I E V R H A - 464

SMC6\_YEAST - I Q N K K V E L Q N I A K G N D T F L M N F D R N M D R L L R T I E - - Q R K N E F E T P A I G P L G S L V T I R K G F - 546  
SMC6\_SCHPO - I S D Q S E H L K V L E D V Q R D K V S A F G K N M P Q L L K L I T - - - R E T R F Q H P P K G P M G K Y M T V K E - - 561  
SMC6\_TAKRU - I D A N R R Q L Q S M E S S R N R L Q R F G D Q M P A L L A A I D E A H K G Q F K H R P R G P L G Y L I S L K D - - 523  
SMC6\_XENLA - L E Q H K R Q L K E L H E S K T D R L K R F G G N M P A L L A A I D E A D K L G R F R K K P V G P L G A C I H L K D - - 555  
SMC6\_HUMAN - L S Y N Q R Q L K E L K D S K T D R L K R F G P N V P A L L E A I D D A Y R Q G H F T Y K P V G P L G A C I H L R D - - 516  
SMC6\_MOUSE - L N Y N Q R Q L K E L K D S K T D R L K R F G P H V P A L L E A I D D A Y R R R Q F T H K P I G P L G A C I H L R D - - 522

SMC6\_YEAST - E K W T R S I Q R A I S S S L N A F V V S N P K D N R L F R D I M R S C G I R S - - N I P I V T Y C L S Q F D Y S K - - 602  
SMC6\_SCHPO - Q K W H L I I E R I L G N V I N G F I V R S H D Q L I L K E L M R Q S N C H A - - - - - T V V V G K Y D P F D Y S S - - 615  
SMC6\_TAKRU - P E L A L S I E I C L K N L V Q A F T C D N Y D D E R V L K S L M T K V L Q H G - R R P A I T T S R F F P K V H D V S V - 582  
SMC6\_XENLA - Q E L A L A V E S C L K G L M F A F C C D N H Q D E R M L Q N I M S R E Y P R G - R R P Q I V N E F I D H V Y D V R Q - 614  
SMC6\_HUMAN - P E L A L A I E S C L K G L L Q A Y C C H N H A D E R V L Q A L M K R F Y L P G T S R P P I I V S E F R N E I Y D V R H - 576  
SMC6\_MOUSE - P E L A L A I E S C L K G L L Q A Y C C H N H A D E R V L Q S L M K K F Y P P G T S R P Q I I V S E F R D E V Y D V R L - 582

SMC6\_YEAST - - G R A H G N Y P T I V D A L E F S K P E I E C L F V D L S R I E R I V L I E D K N E A R N F L Q R N P V N V N M A L - 660  
SMC6\_SCHPO - - G E P D S Q Y P T V L K I I K F D D E V L H T L I N H L G I E K M L L I E D R R E A E A Y M K R G I A N V T Q C Y A - 674  
SMC6\_TAKRU - R A V N H P D Y P S V L Q A L E I D P V V A N C L I D Q R A I E C I L L I K N R T E A R R V M Q G R N P - - - - - 635  
SMC6\_XENLA - R A T F H P D H P T V L T A L E I D H P V T N C L I D M R G I E T I L I I K G K D E A R E I M Q K R A P - - - - - 667  
SMC6\_HUMAN - R A A Y H P D F P T V L T A L E I D N A V V A N S L I D M R G I E T V L L I K N N S V A R A V M Q S Q K P - - - - - 629  
SMC6\_MOUSE - R A A Y H P E F P T V L T A L E I D N A V V A N S L I D M R S I E T V L L I K N N S V A R A V M Q S Q K P - - - - - 635

SMC6\_YEAST - - S L R D R R S G F Q L S G G Y R L D T V T Y Q D K I R L K V - - N S S S O N G T Q Y L K D L I E Q E T K E L Q N I R D - 717  
SMC6\_SCHPO - L D P R N R G Y G F R I V S T Q R S S G I S K V T P W N R P P R I G F S S S T S I E A E K K I L D D L K K Q Y N F A S N - 734  
SMC6\_TAKRU - - P Q N C T S A F S V E G D Q I F T N R S Y T A D Q T R A N F L S K D V E E G I R H L K R E M T Q K V Q A A H I Q - 693  
SMC6\_XENLA - - - P R N C R E A F T G E G D Q V Y T N R Y S S D S R R A T L L S R D V E A E I S H L E K L R N F G S Q M A T F Q Q - 725  
SMC6\_HUMAN - - - P K N C R E A F T A D G D Q V F A G R Y S S E N T R P K F L S R D V D S E I S D L E N E V E N K T A I L N L Q - 687  
SMC6\_MOUSE - - - P K N C R E A F T A D G D Q V F A G R Y S S E S T R P K F L S R D V D S E I S D L E T E I E N K K G H I I T L Q - 693

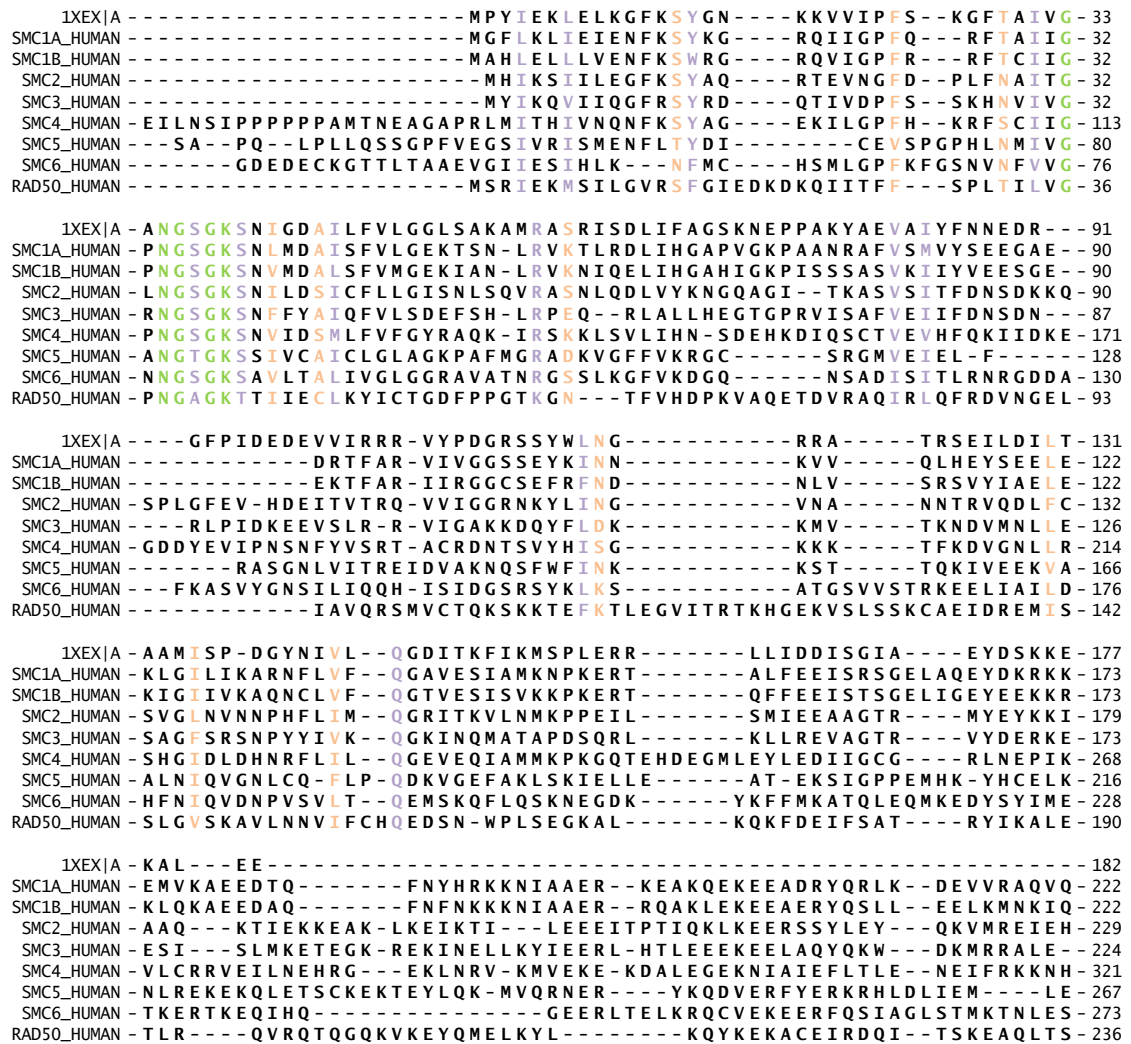
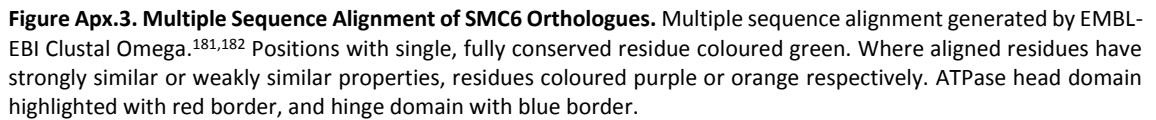
SMC6\_YEAST - R Y E E - - - - - K L S E V R S R L K E - I D G R L K S T K N E M R K T N F R M T E L K M N V G K V V D T G I L N S K - 770  
SMC6\_SCHPO - Q L N E A K I E Q A K F K R D E Q L L V E K I E G I K K R I - - - - - L L K R R E V N S L E S Q E - - L S V - - - 781  
SMC6\_TAKRU - Q I - - - R S T D K N I S E N Q D L L R - T Q T E Q K T T E V K T M K L Q L E L T D L K N V E E P - - - - - 739  
SMC6\_XENLA - R A - - - Q S V D K D I K E N E G I L R Q - Y H N S K K Q I Q I D L R P L L E R I S E L E N V E E Q - - - - - 771  
SMC6\_HUMAN - H L - - - S A L E K D I K H N E E L L K R - C Q L H Y K E L K M K I R K N I S E I R E L E N I E E H - - - - - 733  
SMC6\_MOUSE - R L - - - S A L E K D I K R N E E L L K R - C Q L H Y K E I K M K I R K N I S E I R E L E N I E E H - - - - - 739

SMC6\_YEAST - I N E R K N Q E Q A I A S Y E A A K E E L G L K I E Q I A Q E A Q P I K E - - - - - Q Y D S T K - - - - 813  
SMC6\_SCHPO - - - - - L D T E K I Q T L E R R I S E T E K E L E S Y A G Q L Q D A K N E E H R I - - - R D N Q R P V I E E I R I Y R - 833  
SMC6\_TAKRU - - - - - Q S E D L A E L K S A F E N A E Q E Y K Q H K Q L I D T A A E E - A D V K K V E T K T L P L F T S I K - - - 789  
SMC6\_XENLA - - - - - P S I D I A T L E G A E E N L N K I L Q K V E F L A K E K M G N L K S F L T T A E I N Y E E I S - - - 822  
SMC6\_HUMAN - - - - - Q S V D I A T L E D E A Q E N K S K M K M V E E H M E Q Q K E N M E H L K S L K I E A E N K Y D A I K - - - 784  
SMC6\_MOUSE - - - - - Q S V D I A T L E D E A E E N K I K M Q M V E K N M E Q Q K E N M E N L K S L K I E A E N K Y D T I K - - - 790

SMC6\_YEAST - L A L V E A Q D E L Q Q L K E D I N S R Q S K I Q K Y K D D T I Y Y E D K K K V Y L E N I K K I E V N V - - - A A L - - - 868  
SMC6\_SCHPO - E K I Q T E T Q R L S S L Q T E L S R L R D E K - - - - - R N S E V D I E R H R Q T S I S C T N I L R E K E A K V - 886  
SMC6\_TAKRU - C N I L W V H C L F V L L Q E E L S K T D Q E V M K C K H H E K Y E E R R N A H L C S I V S T L E N N V - - - A S K - - 844  
SMC6\_XENLA - K K I S S V A E V A E P V K E D L H R V D Q E V E N C K R H R K H Y E E K L K E H L D R I Q K R K E E V - - - A A K - - - 877  
SMC6\_HUMAN - F K I N Q L S E L A D P L K D E L N L A D S E V D N Q K R G K R H Y E E K Q K E H L D T L N K K K R E L - - - D M K - - - 839  
SMC6\_MOUSE - L K I N Q L S E L A D P L K D E L N L A D S E V D S Q R G K G K H Y E D K Q K E H L D T L N K K R E L - - - D M K - - - 845

SMC6\_YEAST - - - K E G I Q R Q I Q N A C A F C S K E R I E N V D L P D T Q E E I K R E L D K V S R M I Q K A E K S L G L S Q E V I - 926  
SMC6\_SCHPO - Q C A Q V V A D Y T A K A N T R - C - - - - E R V P V Q L S P A E L D N E I E R L Q M Q I A E W K N R T G V S V E Q A A - 941  
SMC6\_TAKRU - - - E K E L Q E S I A K A K E I C P - - - - - E Q L V V R R T A R S L D V E I T R L K V K I A T Q R E H Q G D R E - E I V - 897  
SMC6\_XENLA - - - E Q E L E V K I S Q A K C I C P - - - - - E R I E V S R T A R S L D T E I N R L R E K I N S E E V L H G N R E - E I I - 930  
SMC6\_HUMAN - - - E K E L E E K M S Q A R Q I C P - - - - - E R I E V E K S A S I L D K E I N R L R Q K I Q A E H A S H G D R E - E I M - 892  
SMC6\_MOUSE - - - E K E L Q E K M S Q A R Q I C P - - - - - E R I E V K K S A S I L D K E I N R L R Q K I Q A E H A S H G D R E - E I M - 898

SMC6\_YEAST - A L F E K C R N K Y K E G Q K K Y M E I D E A L N R L H N S L K A R D Q N Y K N A E K G T C F D A D M D F R A S L K V R - 986  
SMC6\_SCHPO - E D Y L N A K E K H D Q A K V L V A R L T Q L L Q A L E E T L R R R N E M W T K F R K L I T L R T K E L F E L Y L S Q R - 1001  
SMC6\_TAKRU - R E Y H E A L E S Y A N K A Q Q I K N L N F I K L C D R V M D Q R L Y A Y T V L R R F L S A R C K Y Y F D S M L A Q R - 957  
SMC6\_XENLA - K Q Y H E A K E R Y Q D V E G K V K H L K R F I K L L D E I M A Q R Y K S Y Q Q F R R C L T F R C K I Y F D S L L S Q R - 990  
SMC6\_HUMAN - R Q Y Q E A R E T Y L D L D S K V R T L K K F I K L L G E I M E H R F K T Y Q Q F R R C L T L R C K L Y F D N L L S Q R - 952  
SMC6\_MOUSE - K Q Y Q E A R E T Y L D L D N K V R T L R R F I K L L E I M T H R Y K T Y Q Q F R R C L T L R C K L Y F D N L L S Q R - 958



```

1XEX|B -----EKEKKNVFMRTFEAISRN----FSEIFAKLSPGGSSARLI--LENPEDPFS-----44
SMC1A_HUMAN -FEQIKKERFDRFNACFESVATN----IDEIYKALSRNSSAQAF LGPENPEEPYL-----1108
SMC1B_HUMAN -FEQVKRRYDLFTQCFEHSIS----IDQIYKKLCRNNSAQAF LSPENPEEPYL-----1104
SMC2_HUMAN -IEDLDQKKNQALNIAWQKV NKD----FGSIFSTLLPGANAMLA--PPEGGQTV-----1064
SMC3_HUMAN -MNVLELRKYEAQLTFKQVSKN----FSEVFQKLVPGGKATLV--MKKGDVEGSQSQDEG-1071
SMC4_HUMAN -YEDLRKQRLNEFMAGFYITNK----LKENYQMLTLGGDAELE--LVDSLDPPFS-----1171
SMC5_HUMAN -ISQVKERWLNPLKELVEKINEK----FSNFFSSMQCAGEVDLH--TENEEEDYD-----967
SMC6_HUMAN -LGEIMEHRFKTYQQFRRCLTLRCKLYFDNLLSQRAYCGKMNFD--HKNE-----965
RAD50_HUMAN -YKTL DQAIMKFHSMKMEENKI---IRDLWRSTYRGQDIEYIEIRS DADENVS-----1174

1XEX|B -----GGLEIEAKPAGK---DVKRIEAMSGGEKALTALAFV-77
SMC1A_HUMAN -----DGINYN CVAPGK---RFRPMDNLSGGEKTVAAALALL-1141
SMC1B_HUMAN -----EGISYN CVAPGK---RFMPMDNLSGGEKCVAAALALL-1137
SMC2_HUMAN -----LDGLEFKVALGNT--WKENLTELGGQKSLVALSLI-1098
SMC3_HUMAN -EGSGESERGSQSSVPSVDQFTGVGIRVSFTGKQG---EMREMQLSGGQKSLVALALI-1128
SMC4_HUMAN -----EGIMFSVRPPKK---SWKKIFNLGGGEKTLSSLALV-1204
SMC5_HUMAN -----KYGIRIRVKFRSSTQLHELTPHHQSGGERSYSTMLYL-1004
SMC6_HUMAN -----TLSISVQPGEGNKAAFNMDMRALSGGERSFSTVCFI-1000
RAD50_HUMAN -----ASDKRRNYNRYVVMLKGDT---ALDMRGRCAGQKVLASLIIR-1214

1XEX|B -FA--IQKFKPAPFYLFDQIDAHLDDANVKRVADLIKESSK-----ESQFIVITLRDVM-129
SMC1A_HUMAN -FA--IHSYKPAPFVFLDEIDAALDNTNIGKVANYIKEQST-----CNFQAIVISLKEEFY-1194
SMC1B_HUMAN -FA--VHSFRPAPFVFLDEVDAALDNTNIGKVSSYIKEQTQ-----DQFQMIVISLKEEFY-1190
SMC2_HUMAN -LS--MLLFKPAPTYILDEVDAALDLSHTQNIQGM LRTHTF-----HSQFIVVSLKEGMF-1150
SMC3_HUMAN -FA--IQKCDPAPFYLFDQIDQALDAQHRKAVSDMIMELAV-----HAQFITTFRPELL-1180
SMC4_HUMAN -FA--LHHYKPTPLYFMDIDAALDFKNVSIVAFYIYEQTK-----NAQFIISLRNNMF-1256
SMC5_HUMAN -MA--LQELNRCPRFVVD E INQGM D PINERRVFEMVNTAC---KENTSQYFFITPK--LL-1057
SMC6_HUMAN -LS--LWSIAESPFRCLDFDVYMDMVNRRIAMDL LK MAD--SQRFQFILLTPQ--SM-1053
RAD50_HUMAN -LALAETFCLNCGTIALDEPTTNLDRENIESLAHALVEIISKRSQQRNFL L L V I T H D E D F V -1274

1XEX|B -ANA-----DKIIGVSMRDGVSKVVSLSL-EKAMKILEEIRKKQGW EHG N-----172
SMC1A_HUMAN -TKA-----ESLIGVYPEQGD CVISKVLTFDLT-KYPDANPNPNEQ-----1233
SMC1B_HUMAN -SRA-----DALIGIYPEYDDCMFSRVLTLDLS-QYPDTEGQESSKRHG ESR-----1235
SMC2_HUMAN -NNA-----NVLFKTKFVDGVSTVARFTQCQNGK-ISKEAKSKAKPPKGAHVEV-----1197
SMC3_HUMAN -ESA-----DKFYGVKFRNKVSHIDVITA-EMAKDFVEDDTT-----HG-----1217
SMC4_HUMAN -EIS-----DRLIGIYKTYNITKSVAVNPKEIASKGLC-----1288
SMC5_HUMAN -QNL PY---SEKMTVLFVYNGPHMLEPNT-----WNLKAFQRRRRRITFTQPS-1101
SMC6_HUMAN -SSLPS---SKLIRILRMSDP-----ER-----GQTTLPFRPVTQEEDDDQR-1091
RAD50_HUMAN -ELLGRSEYVEKFYRIKKNIDQCS--EIVKCSVSSLGFNVH-----1312

```

**Figure Apx.4. Multiple Sequence Alignment of SMC Human Homologues with Template Structure (PDB ID: 1XEX).<sup>224</sup>**

Multiple sequence alignment generated by EMBL-EBI Clustal Omega.<sup>181,182</sup> Positions with single, fully conserved residue coloured green. Where aligned residues have strongly similar or weakly similar properties, residues coloured purple or orange respectively. ATPase head domain highlighted with red border, and hinge domain with blue border.



## VII.II. Histone Acetyltransferase Domain Multiple Sequence Alignment

KAT2A_HUMAN	-VMLTITDPAA---MLGPETSLLSANAARDETARLEERRGIIIEFHVIGNSLT-----	509
KAT2B_HUMAN	-VMSTITDPAA---MLGPETNFLSAHSARDEAARLEERRGVIEFHVVGNSLN-----	504
ELP3_HUMAN	-GVEHGNLREL---ALARMKD-LG-IQCRDVR---TREVGIIQETHHKVRPYQVELVRRDYV-	428
KAT5_HUMAN	-E--DSQDSSDGI PSAPRMTGSLVSDRSHDDIVTRMKNIECIEL-----	237
KAT6A_HUMAN	-EIMTEKDMELFRDIQE---QALQKVGVGTGPPDPQVRCPSVIEF-----	514
KAT6B_HUMAN	-DVTTEEDLDVFKQAQE---LSWEKIECESGVEDCGRYPSVIEF-----	725
KAT7_HUMAN	-----YDLDLFRRARASEDLEKLRLQGQITEGSNMIKTIAF-----	342
KAT8_HUMAN	-----MDPTTAALKEHEAITKVKYVDKIHT-----	184
KAT2A_HUMAN	-----PKANRRVLLWLVLGLQNVFSHQLPRMPKEYIARLVFDPKHKTALAIKDG-	557
KAT2B_HUMAN	-----QKPNKKILMWLVGLQNVFSHQLPRMPKEYITRLVFDPKHKTALAIKDG-	552
ELP3_HUMAN	-ANGGWETFLSYEDPDQDILIGLLRLRKCE-----ETFRFELG-GGVSIVRLE-	475
KAT5_HUMAN	--GRH-----RLKPWYF---SPYPQELTTLPVLYLCEFCFLKYGRSLKCLQRHLLT-	281
KAT6A_HUMAN	--GKY-----EHTWYS---SPYPQEYSRLPKLYLCEFCFLKYMKSRITLQQHMK-	558
KAT6B_HUMAN	--GKY-----EIQTWYS---SPYPQEYARLPKLYLCEFCFLKYMKSKNILLRHSK-	769
KAT7_HUMAN	--GRY-----ELDTWYH---SPYPEEYARLGRLYMCEFCFLKYMKSQITILRRHMA-	386
KAT8_HUMAN	--GNY-----ETDAWYF---SPFPEDYGKQPKLWLCEYCLKYMKYKESYRFLHG-	228
KAT2A_HUMAN	-----RVIGGICFRMFPTQGFTIEIVFCAVTSNEQVKGYGTHLMNHLKEYHIKH-N-	606
KAT2B_HUMAN	-----RVIGGICFRMFPSQGFTIEIVFCAVTSNEQVKGYGTHLMNHLKEYHIKH-D-	601
ELP3_HUMAN	-----HVYGSV-----VPVSSRDPTKFQHQGFGLMLEEAERIAREHFG-	514
KAT5_HUMAN	-KCDLRHPPGNEIYRKGTISFFIEIDGRKNKSYSQNL-----LLAKCFLDHKTLYYDTPPF-	336
KAT6A_HUMAN	-KCGWFHPPANEIYRKNNISVFVDGNVSTIYCNLC-----LLAKLFLDHKTLYYDVEPF-	613
KAT6B_HUMAN	-KCGWFHPPANEIYRRKDLVFEVDGNMMSKIYCNLC-----LLAKLFLDHKTLYYDVEPF-	824
KAT7_HUMAN	-KCVWKHPPGDEIYRKGSISVFVDGKKNKIYCNLC-----LLAKLFLDHKTLYYDVEPF-	441
KAT8_HUMAN	-QCQWRQPPGKEIYRKSNISVFEVDGKDHKIYCNLC-----LLAKLFLDHKTLYYDVEPF-	283
KAT2A_HUMAN	-ILYFLTYA---DEYAIQGYFKKQGFSDK-----IKVPKSRYLGYIKD-----YEGATLM-	651
KAT2B_HUMAN	-ILNFLTIA---DEYAIQGYFKKQGFSE-----IKIPKTKYVGYIKD-----YEGATLM-	646
ELP3_HUMAN	-SGKIAVIS---GVGTRNYRKIGY-----	535
KAT5_HUMAN	-LFYVMT EYDCKGFHIVGYFSKEKESTEDYNVACILTLPYPYQRRGYGKLLIEFSYELSKVE-	396
KAT6A_HUMAN	-LFYVLTQNDVKGCHLVGYFSKEKHCQKYNVSCIMILPQYQKGYGRFLIDFSYLLSKRE-	673
KAT6B_HUMAN	-LFYVLTKNDEKCHLVGYFSKEKLCQKYNVSCIMIMPQHQRQGFGRFLIDFSYLLSKRE-	884
KAT7_HUMAN	-LFYVMT EADNTGCHLIGYFSKEKNSFLNYNVSCILTMPQYMRQGYGKMLIDFSYLLSKVE-	501
KAT8_HUMAN	-VFYILT EVDREQAHIVGYFSKEKESPDGNNVACILTLPYPYQRRGYGKFLIAFSYELSKLE-	343
KAT2A_HUMAN	-ECELNPRIPYTELSHIKK--QKE-IKKLIERKQAQIRKVYPGLSCFKEGVRQIPVESV-	708
KAT2B_HUMAN	-GCELNPRIPYTEFSVIKK--QKE-IKKLIERKQAQIRKVYPGLSCFKDGVRRQIPIESI-	703
ELP3_HUMAN	-----	535
KAT5_HUMAN	-GKTGTPEKPLSDLGLLSYRSYWSQTILEILMG-----LKSESGERPQITINEI-	444
KAT6A_HUMAN	-GQAGSPEKPLSDLGRLSYMAYWKSIVILECLYH-----QND-----KQISIKKL-	716
KAT6B_HUMAN	-GQAGSPEKPLSDLGRLSYLAYWKSIVILEYLYH-----HHE-----RHISIKAI-	927
KAT7_HUMAN	-EKVGSPEKPLSDLGLLSYRSYWKEVLLRYLHN-----FQG-----KEISIKAI-	544
KAT8_HUMAN	-STVGSPEKPLSDLGKLSYRSYWSVLLLEILRD-----FRG-----T-LSIKDL-	385

**Figure Apx.5. Multiple Sequence Alignment of GNAT and MYST Family HAT Domains.** Multiple sequence alignment generated by EMBL-EBI Clustal Omega.<sup>181,182</sup> Positions with single, fully conserved residue coloured green. Where aligned residues have strongly similar or weakly similar properties, residues coloured purple or orange respectively. HAT domain highlighted with red border.

## VII.III. Bromodomain Multiple Sequence Alignments

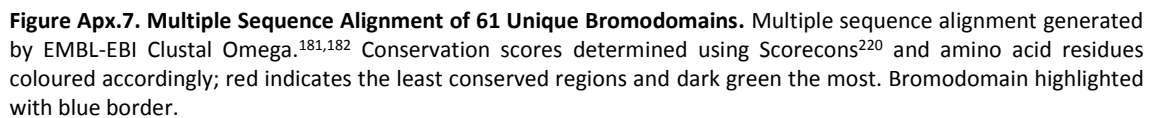
KAT2A_HUMAN	-----EKGKELKD--PDQLYTTLKNLLAQI	KSHPSAWPFMEPVKKS--EAPDYY	766
KAT2B_HUMAN	-----EKSKEPRD--PDQLYSTLKSILQQV	KSHQSAPWFMEPVKRT--EAPGYY	761
BPTF_HUMAN	-----EDAMTVLTPLEKDYEGLRVLRSL	QAHKMAWPFLEVPDPN--DAPDYY	2965
BRDT(1)_HUMAN	-----KKNGRLTN--QLQYLQKVVVKDLWK	--HSFSWPFQRPVDAVKLQLPDYY	67
BRD4(1)_HUMAN	-----NKPQRQTN--QLQYLLRVVVKTLWK	--HQFAWPFQQPVDAVKLNLDPYY	98
BRD3(1)_HUMAN	-----SKPGRKTN--QLQYMQNVVVKTLWK	--HQFAWPFYQPVDAIKLNLDPYH	74
BRD2(1)_HUMAN	-----KKPGRVTN--QLQYLHKVVMKALWK	--HQFAWPFQRPVDAVKLGLPDYH	114
BRDT(2)_HUMAN	-----VKTVKVTE--QLRHCSEILKEMLAK	KHFSYAWPFYNPVDVNALGLHNY	310
BRD4(2)_HUMAN	-----EKSSKVSE--QLKCCSGILKEMFAK	KHAAYAWPFYKPVDEALGLHDYC	391
BRD3(2)_HUMAN	-----GKKGKLSE--HLRYCDSILREMLS	KHAAYAWPFYKPVDAEALGLHDYH	349
BRD2(2)_HUMAN	-----SKKGKLSE--QLKHCHGILKELLS	KHAAYAWPFYKPVDAALGLHDYH	387
PHIP(1)_HUMAN	-TDGECRSLIYKPLDGEWGTNPRDEECERIVAGINQL	MTLDIASAFVAPVDLQ--AYPMYC	1197
BRWD1(1)_HUMAN	-TTDELEKLLYKPQAGEWGQKSRDEECDRIISGIDQL	LNLDIAAFAGPVDLC--TYPKYC	1198
BRWD3(1)_HUMAN	-SQEELTALLYKPQEGEWGAHSRDEECERVIQGINHL	LSLDFASPAFVPDLS--AYPLYC	1179
KAT2A_HUMAN	-EVIREFIDLKMTTERLSRYVYVTRKLFVADLQRVIAN	CREYNPPDSEYRCASALEKFFY	826
KAT2B_HUMAN	-EVIREFMDLKTMSERLKNRYVYVSKKLFMADLQRVFTN	CKEYNPPESEYKCANILEKFFF	821
BPTF_HUMAN	-GVIKEPMDLATMEERVQRYYEKLTEFVADMTKIFDN	CRYNPSDSPFYQCAEVLESFFV	3025
BRDT(1)_HUMAN	-TIIKNPMDLNTIKKRLENKYAKASECIEDFNTMFSN	CYLYNKPGGDIVLMAQALEKLFM	127
BRD4(1)_HUMAN	-KIIKTPMDMGTIKKRLENYYWNAQECIQDNTMFTN	CYIYNKPGGDIVLMAEALEKFL	158
BRD3(1)_HUMAN	-KIIKNPMDMGTIKKRLENYYWSASECMQDNTMFTN	CYIYNKPTDDIVLMAQALEKIFL	134
BRD2(1)_HUMAN	-KIIKQPMDMGTIKKRLENYYWAASECMQDNTMFTN	CYIYNKPTDDIVLMAQTLKIFL	174
BRDT(2)_HUMAN	-DVVKNPMDLGTIKEKMDNQEKDAYKFAADVRLFMN	CYKYNPPDHEVVTMARMLQDVFE	370
BRD4(2)_HUMAN	-DIIKHPMDMSTIKSKLEAREYRDAQEFGADVRLMFS	CYKYNPPDHEVVMARKLQDVFE	451
BRD3(2)_HUMAN	-DIIKHPMDLSTVKKRMDGREYPDAGGFAADVRLMFS	CYKYNPPDHEVVMARKLQDVFE	409
BRD2(2)_HUMAN	-DIIKHPMDLSTVKKRMENRDYRDAQEFAADVRLMFS	CYKYNPPDHDVVMARKLQDVFE	447
PHIP(1)_HUMAN	-TVVAYPTDLSIIKQRLLENFYRRVSSLMWEVRYIEHN	TRTFNEPGSPIVKSAKFVTDLLL	1257
BRWD1(1)_HUMAN	-TVVAYPTDLYTIRMRLVNRFYRRLSALVWEVRYIEHN	ARTFNEPESVIARSAKKITDQLL	1258
BRWD3(1)_HUMAN	-TVVAYPTDLNTIRRLLENRFYRRISALMWEVRYIEHN	ARTFNEPDSPIVKAAKIVTDVLL	1239

**Figure Apx.6. Multiple Sequence Alignment of Type I and Type II Bromodomains.** Multiple sequence alignment generated by EMBL-EBI Clustal Omega.<sup>181,182</sup> Positions with single, fully conserved residue coloured green. Where aligned residues have strongly similar or weakly similar properties, residues coloured purple or orange respectively. Bromodomain highlighted with blue border.

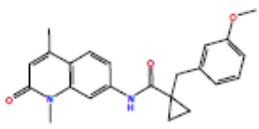
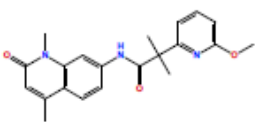
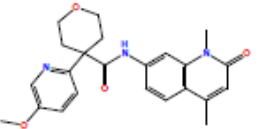
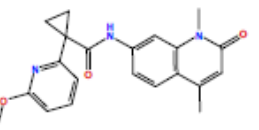
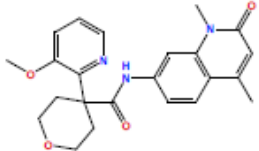
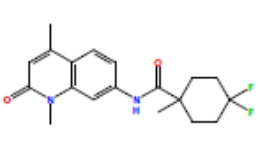
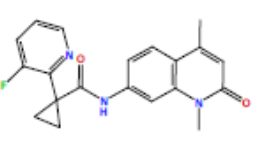
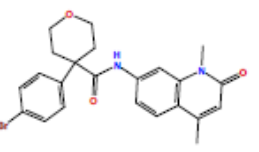
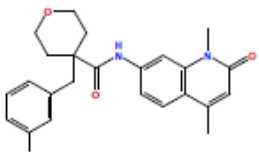
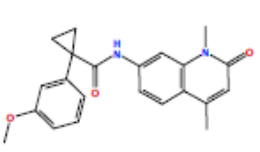
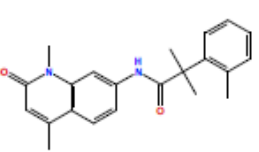
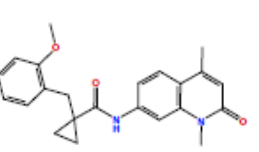
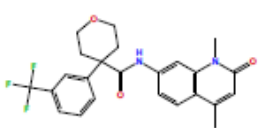
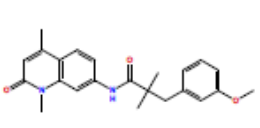
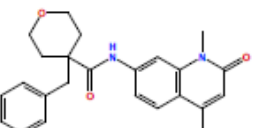
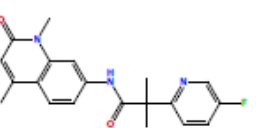
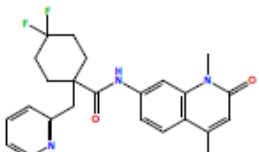
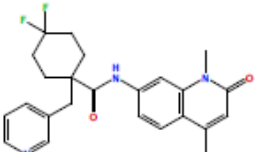
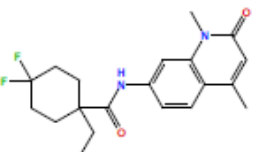
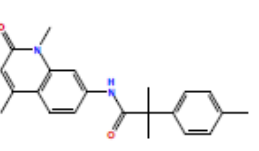
KAT2A_HUMAN	-YTTLKNLLAQIK	----	SHPSAWPFMEPVKKSS	----	EAP	-	763																																																																																																																																																																																																																																																																																																																																																																																																		
KAT2B_HUMAN	-YSTLKSILQQVK	----	SHQSAWPFMEPVKRT	----	EAP	-	758																																																																																																																																																																																																																																																																																																																																																																																																		
BPTF_HUMAN	-EKDYEGLKRVLR	SLQA	AHKMAWPFLEVPDNP	----	D--AP	-	2962																																																																																																																																																																																																																																																																																																																																																																																																		
BRWD3(2)_HUMAN	-KKQCKELLSLIY	----	EREDSEPFRRQPADLL	----	S--YP	GHQEQEGESSESVP	1356																																																																																																																																																																																																																																																																																																																																																																																																		
BRWD1(2)_HUMAN	-KKQCKELVNLI	F----	QCEDSEPFRRQVDLV	----	E--YP	-	1348																																																																																																																																																																																																																																																																																																																																																																																																		
PHIP(2)_HUMAN	-KKQCEELLNLI	F----	QCEDSEPFRRQVDLL	----	E--YP	-	1351																																																																																																																																																																																																																																																																																																																																																																																																		
CECR2_HUMAN	-DDDFAMYKVL	DVVK	AHKDSWPFLEPVDES	----	Y--AP	-	469																																																																																																																																																																																																																																																																																																																																																																																																		
BRDT(1)_HUMAN	----	YLKQVVVKLD	LWK--HSFSWPFRRQPVDAV	----	KLQLP	-	64																																																																																																																																																																																																																																																																																																																																																																																																		
BRD4(1)_HUMAN	----	YLLRVVLKTK	LWK--HQFAWPFRRQPVDAV	----	KLNL	P	95																																																																																																																																																																																																																																																																																																																																																																																																		
BRD3(1)_HUMAN	----	YMQNVVVKT	LWK--HQFAWPFYQVDAI	----	KLNL	P	71																																																																																																																																																																																																																																																																																																																																																																																																		
BRD2(1)_HUMAN	----	YLHKVVMKA	LWK--HQFAWPFRRQPVDAV	----	KLGLP	-	111																																																																																																																																																																																																																																																																																																																																																																																																		
BRDT(2)_HUMAN	----	HCSEILKEM	LAKKHFSYAWPFYNPVDVN	----	ALGLH	-	307																																																																																																																																																																																																																																																																																																																																																																																																		
BRD4(2)_HUMAN	----	CCSGILKEM	FAKKHAAYAWPFYKPVDE	----	ALGLH	-	388																																																																																																																																																																																																																																																																																																																																																																																																		
BRD3(2)_HUMAN	----	YCDSILREM	LSKKHAAYAWPFYKPVDAE	----	ALGLH	-	346																																																																																																																																																																																																																																																																																																																																																																																																		
BRD2(2)_HUMAN	----	HCNGILKEL	LSKKHAAYAWPFYKPVDAE	----	ALGLH	-	384																																																																																																																																																																																																																																																																																																																																																																																																		
TRIM66_HUMAN	-QKKCKEKLVL	SLC----	CNNLSLPFHEPVSP	----	L--AR	-	1075																																																																																																																																																																																																																																																																																																																																																																																																		
TRIM28_HUMAN	-QKCKERVLLAL	F----	CHEPCRPLHQLATDS	----	T--T	-	728																																																																																																																																																																																																																																																																																																																																																																																																		
TRIM24_HUMAN	-KKRKCERLLFLY	----	CHEMSLAFQDPVPLT	----	--VP	-	933																																																																																																																																																																																																																																																																																																																																																																																																		
TRIM33-1_HUMAN	-QKRKCERLLLY	----	CHELSIEFQEPVPAS	----	--IP	-	991																																																																																																																																																																																																																																																																																																																																																																																																		
TRIM33-2_HUMAN	-QKRKCERLLLY	----	CHELSIEFQEPVPAS	----	--IP	-	991																																																																																																																																																																																																																																																																																																																																																																																																		
BAZ2A_HUMAN	-LTFCEIILMEME	----	SHDAAWPFLEPVNPR	----	--LVS	-	1828																																																																																																																																																																																																																																																																																																																																																																																																		
BAZ2B_HUMAN	-LALCSMILTEME	----	THEDAWPFLLPVNLK	----	--LVP	-	2095																																																																																																																																																																																																																																																																																																																																																																																																		
SP110-1_HUMAN	----	----	----	----	I--RDYG	-	610																																																																																																																																																																																																																																																																																																																																																																																																		
SP110-6_HUMAN	----	KCEFLLLKAY	CH--PQSSFFTGPFPN	----	I--RDYG	-	634																																																																																																																																																																																																																																																																																																																																																																																																		
SP100_HUMAN	----	KKKFKDP--	----	NAPKRPLAFF	-LFCSEYR	-	784																																																																																																																																																																																																																																																																																																																																																																																																		
SP140_HUMAN	----	KCEFLLLKVY	CC--SESSF	FAKIPYY	YYIR--	-	789																																																																																																																																																																																																																																																																																																																																																																																																		
SP140L_HUMAN	----	KCEFLLLKVY	CC--SESSF	FAKIPYY	YYIR--	-	502																																																																																																																																																																																																																																																																																																																																																																																																		
BRD8(1)_HUMAN	-KKAIMLVWRAA	----	NHRYANVFLQPVTD	----	--IAP	-	742																																																																																																																																																																																																																																																																																																																																																																																																		
BRD8(2)_HUMAN	-KKTLLPVWKMI	A----	SHRFSSPFLKPVSR	----	--QAP	-	1138																																																																																																																																																																																																																																																																																																																																																																																																		
MLL_HUMAN	----	TEESIPSRSS	----	PEGDPPVLTEV	--	-	1697																																																																																																																																																																																																																																																																																																																																																																																																		
PB1(6)_HUMAN	-HNLFVSVMSH-	----	QDDEGRCYSDSLAEI	----	--PAV	-	810																																																																																																																																																																																																																																																																																																																																																																																																		
BRD9_HUMAN	-QQLLEHFLRQLQ	----	RKDPHGFFAFPVDA	----	--IAP	-	171																																																																																																																																																																																																																																																																																																																																																																																																		
BRD7_HUMAN	-QEALNQLMRQLQ	----	RKDPSAFFSPVTD	----	--IAP	-	166																																																																																																																																																																																																																																																																																																																																																																																																		
TAF1(1)_HUMAN	-SSILESIIINDMR	----	DLPNTYFPHTPVNAK	----	--VVK	-	1415																																																																																																																																																																																																																																																																																																																																																																																																		
TAF1L(1)_HUMAN	-SSILESIIINDMR	----	DLPNTHFPHTPVNAK	----	--VVK	-	1434																																																																																																																																																																																																																																																																																																																																																																																																		
BAZ1B_HUMAN	-LQKCEEILHKIV	----	KYRFSWPFREPVTDR	----	--EAE	-	1374																																																																																																																																																																																																																																																																																																																																																																																																		
ATAD2_HUMAN	-RIFLRNVTHRLA	----	IDKRFRVFTKVPDP	----	--EVP	-	1019																																																																																																																																																																																																																																																																																																																																																																																																		
ATAD2B_HUMAN	-RLFRLRDVTKRLA	----	TDKRFNIFSKPVDI	----	--EVS	-	993																																																																																																																																																																																																																																																																																																																																																																																																		
ZMYND11_HUMAN	-GTYLRRFIVSRMK	----	E-----RAIDLN	KKGKDNKHP	-	-	186																																																																																																																																																																																																																																																																																																																																																																																																		
ZMYND8_HUMAN	-SYLLKFAIQKM	----	KQPGTDAFQKVPLE	----	--QHP	-	183																																																																																																																																																																																																																																																																																																																																																																																																		
PHIP(1)_HUMAN	----	RIVAGINQLM	----	TLDIASAFVAVDLQ	----	--AYP	1194																																																																																																																																																																																																																																																																																																																																																																																																		
BRWD1(1)_HUMAN	----	RIISGIDQLL	----	NLDIAAAFAGPVDLC	----	--TYP	1195																																																																																																																																																																																																																																																																																																																																																																																																		
BRWD3(1)_HUMAN	----	RVIQGINHL	----	SLDFASPF	FAVPDLS	----	--AYP	1176																																																																																																																																																																																																																																																																																																																																																																																																	
CREBBP_HUMAN	-QALMPTLEALYR	----	QDPESLPFRQVPDPQ	----	--LLGIP	-	1123																																																																																																																																																																																																																																																																																																																																																																																																		
P300_HUMAN	-QALMPTLEALYR	----	QDPESLPFRQVPDPQ	----	--LLGIP	-	1087																																																																																																																																																																																																																																																																																																																																																																																																		
BRPF1-1_HUMAN	-LILLRKLTLEQLQ	----	EKDTGNI	FSEVPVLS	EVTELDEV	-	663																																																																																																																																																																																																																																																																																																																																																																																																		
BRPF1-2_HUMAN	-LILLRKLTLEQLQ	----	EKDTGNI	FSEVPVLS	EVTELDEV	-	669																																																																																																																																																																																																																																																																																																																																																																																																		
BRD1_HUMAN	-TVLLRSVLDQLQ	----	DKDPARIFAQVSLK	----	--EVP	-	597																																																																																																																																																																																																																																																																																																																																																																																																		
BRPF3_HUMAN	-NVLLRTTLDLLQ	----	EKDPAHIFAEPVNL	SL	--EVP	-	624																																																																																																																																																																																																																																																																																																																																																																																																		
ASH1L_HUMAN	----	--DS	--RQALAAP	LLNLPPKK	----	--KNA	2481																																																																																																																																																																																																																																																																																																																																																																																																		
TAF1(2)_HUMAN	-FILDNIIVTKMM	----	AVPDSWPFHHPVNKK	----	--FVP	-	1538																																																																																																																																																																																																																																																																																																																																																																																																		
TAF1L(2)_HUMAN	-FILDNIIVTKMM	----	AVPDSWPFHHPVNKK	----	--FVP	-	1557																																																																																																																																																																																																																																																																																																																																																																																																		
BAZ1A_HUMAN	----	--QLVVELV	----	RHDDSWPF	FLKLVSKI	----	--QVP	1464																																																																																																																																																																																																																																																																																																																																																																																																	
PB1(3)_HUMAN	----	----	----	IAEPFYHL	PLPSKK	----	--KYP	418																																																																																																																																																																																																																																																																																																																																																																																																	
PB1(2)_HUMAN	----	----	----	ISELFQKL	PLPSKV	----	--QYP	218																																																																																																																																																																																																																																																																																																																																																																																																	
PB1(4)_HUMAN	----	----	----	LCDLFMV	KPSKK	----	--DYP	556																																																																																																																																																																																																																																																																																																																																																																																																	
PB1(1)_HUMAN	----	----	----	LCEL	FI	RAKPKRR	----	--NQP	82																																																																																																																																																																																																																																																																																																																																																																																																
PB1(5)_HUMAN	----	----	----	LSAIFLR	LP	SR	----	--ELP	694																																																																																																																																																																																																																																																																																																																																																																																																
SMARCA4_HUMAN	----	----	----	LSEVFIQL	PSRK	----	--ELP	1495																																																																																																																																																																																																																																																																																																																																																																																																	
SMARCA2-1_HUMAN	----	----	----	LSEVFIQL	PSRK	----	--ELP	1437																																																																																																																																																																																																																																																																																																																																																																																																	
SMARCA2-2_HUMAN	----	----	----	LSEVFIQL	PSRK	----	--ELP	1419																																																																																																																																																																																																																																																																																																																																																																																																	
Conservation	0.0122	0.0228	0.0226	0.0336	0.0323	0.0333	0.0211	0.0101	0.0101	0.0121	0.0216	0.0228	0.0247	0.0282	0.0320	0.0340	0.0363	0.0373	0.0323	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

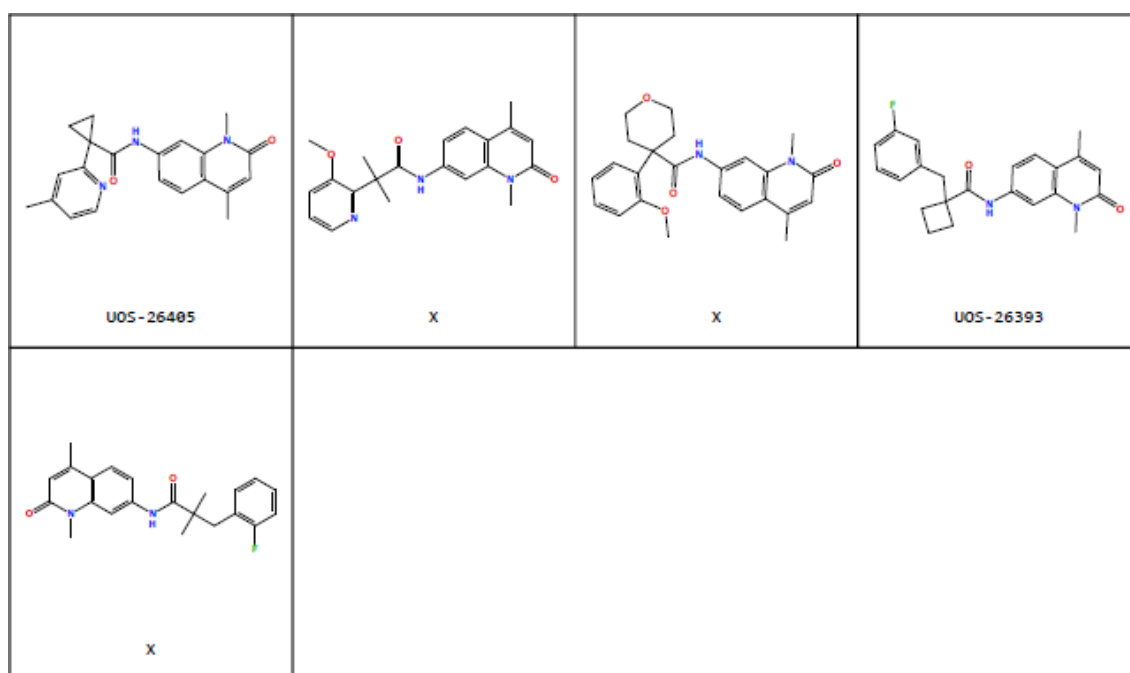


[illegible]

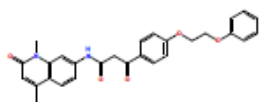
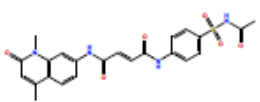
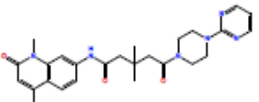
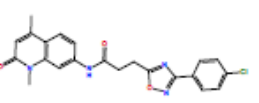
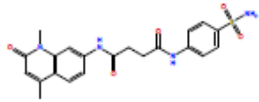
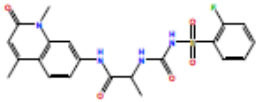
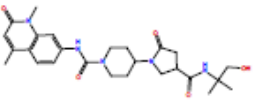
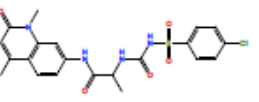
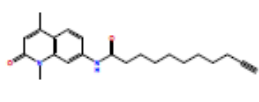
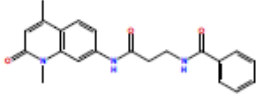
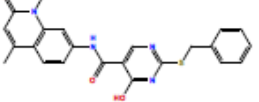
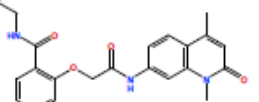
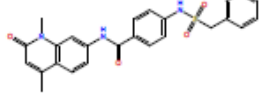
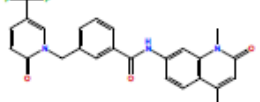
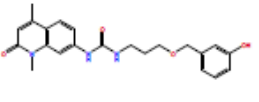
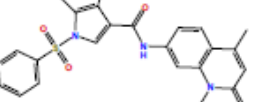
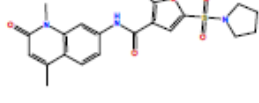
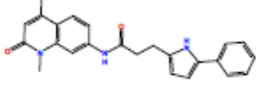
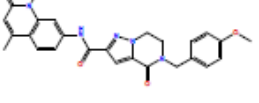
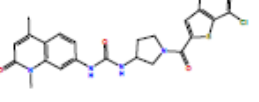


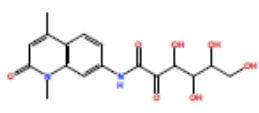
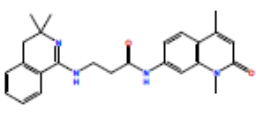
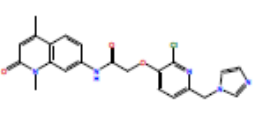
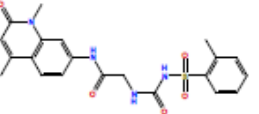
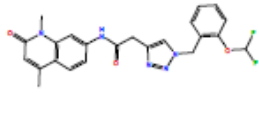
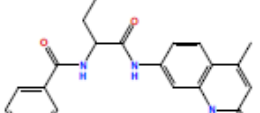
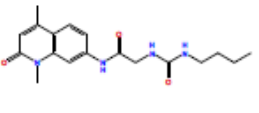
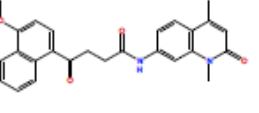
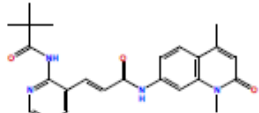
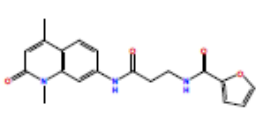
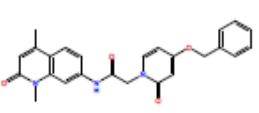
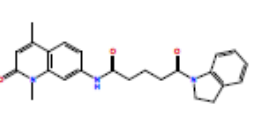
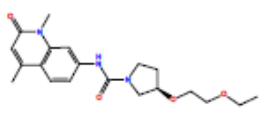
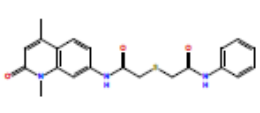
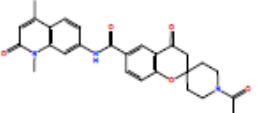
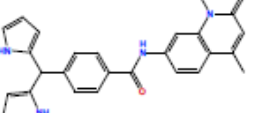
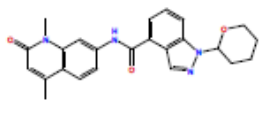
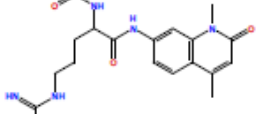
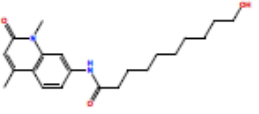
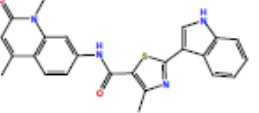
## VII.IV. Quinolinone Screening Libraries

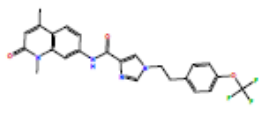
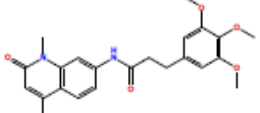
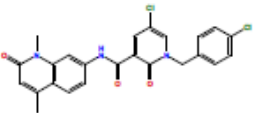
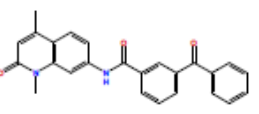
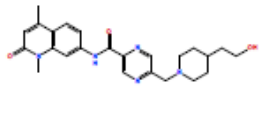
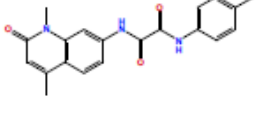
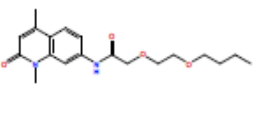
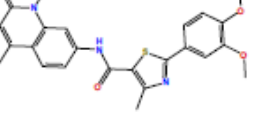
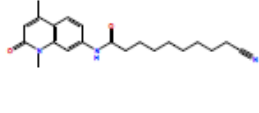
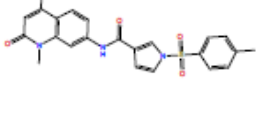
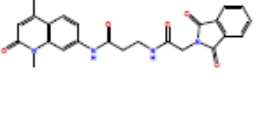
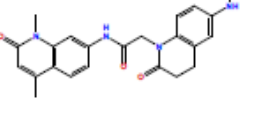
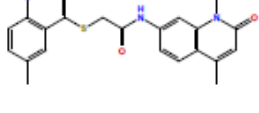
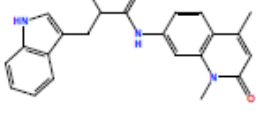
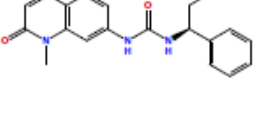
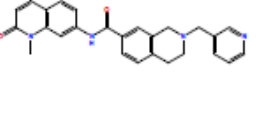
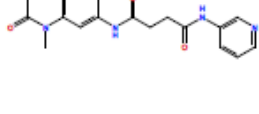
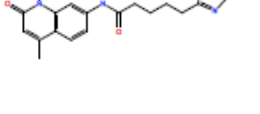
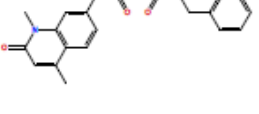
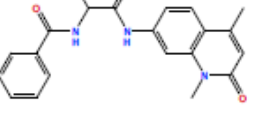
			
U05-26403	U05-26400	U05-26404	U05-26399
			
X	X	U05-26398	X
			
X	U05-26392	X	U05-26406
			
U05-26394	U05-26402	U05-26397	U05-26401
			
U05-26396	X	X	X

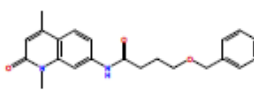
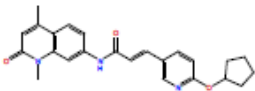
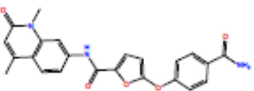
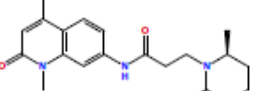
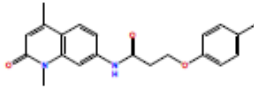
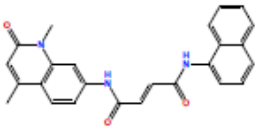
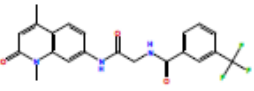
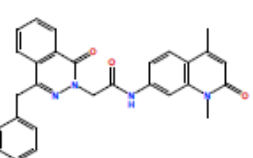
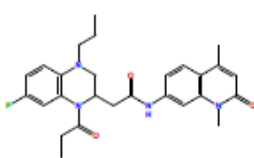
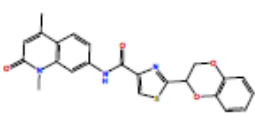
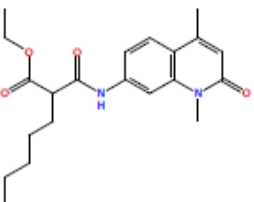
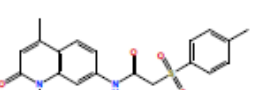
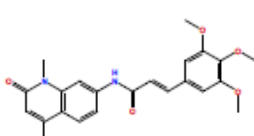
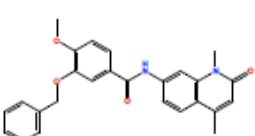
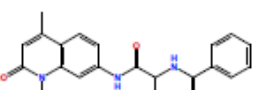
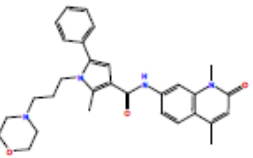
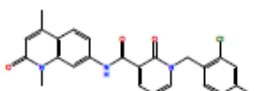
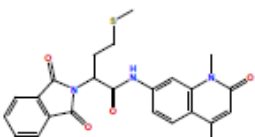
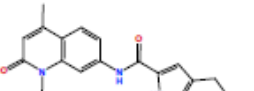
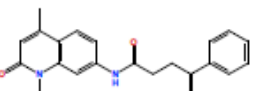


**Figure Apx.8. Amides from Acids Available In-House Selected for Synthesis.** 25 amides selected via docking studies as potential KAT2A and KAT2B inhibitors. Those successfully synthesised labelled with identification codes.

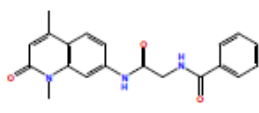
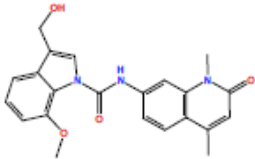
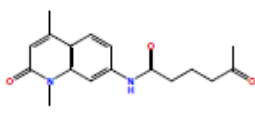
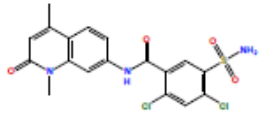
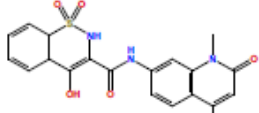
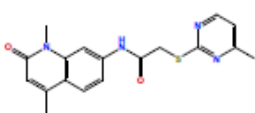
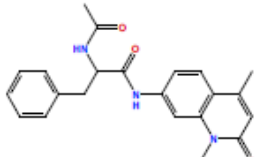
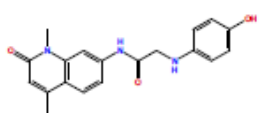
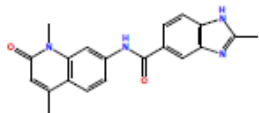
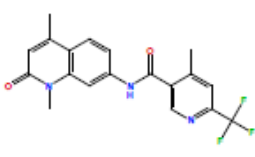
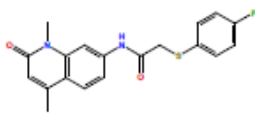
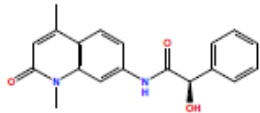
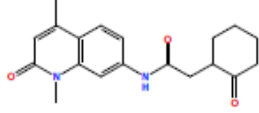
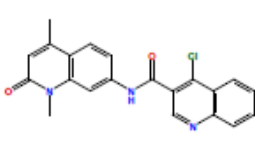
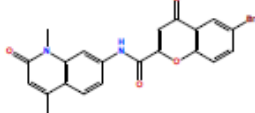
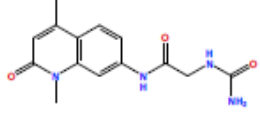
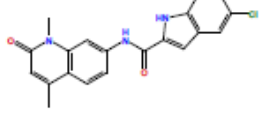
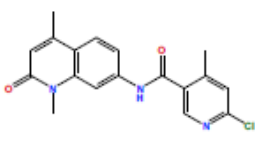
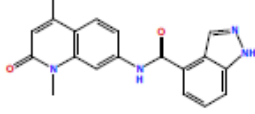
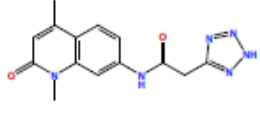
			
X	X	X	U05-26663
			
X	X	X	X
			
X	U05-26653	X	X
			
U05-26660	U05-26678	X	U05-26667
			
X	U05-26685	U05-26676	X

			
X	U05-26680	X	X
			
U05-26692	X	U05-26671	X
			
X	U05-26683	U05-26693	U05-26681
			
X	X	U05-26687	X
			
X	X	U05-26659	X

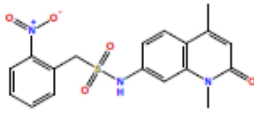
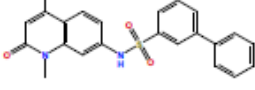
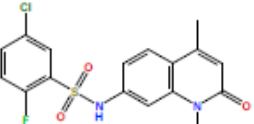
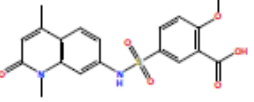
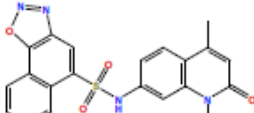
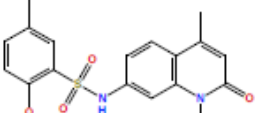
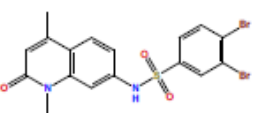
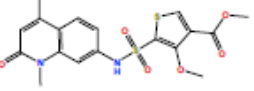
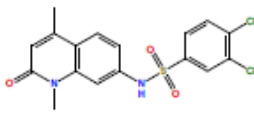
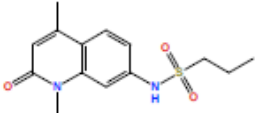
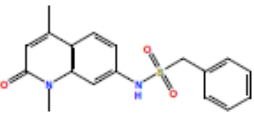
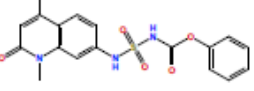
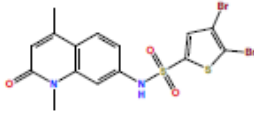
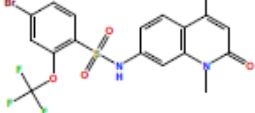
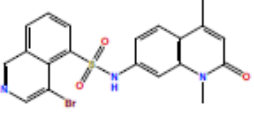
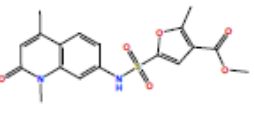
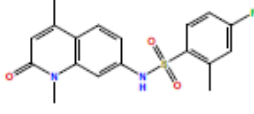
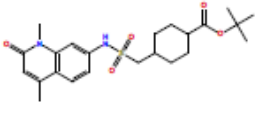
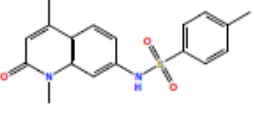
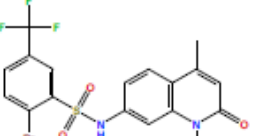
			
UOS-26694	UOS-26654	UOS-26645	UOS-26658
			
X	X	UOS-26684	UOS-26696
			
UOS-26679	UOS-26677	X	UOS-26689
			
UOS-26674	X	X	UOS-26688
			
X	UOS-26682	X	X

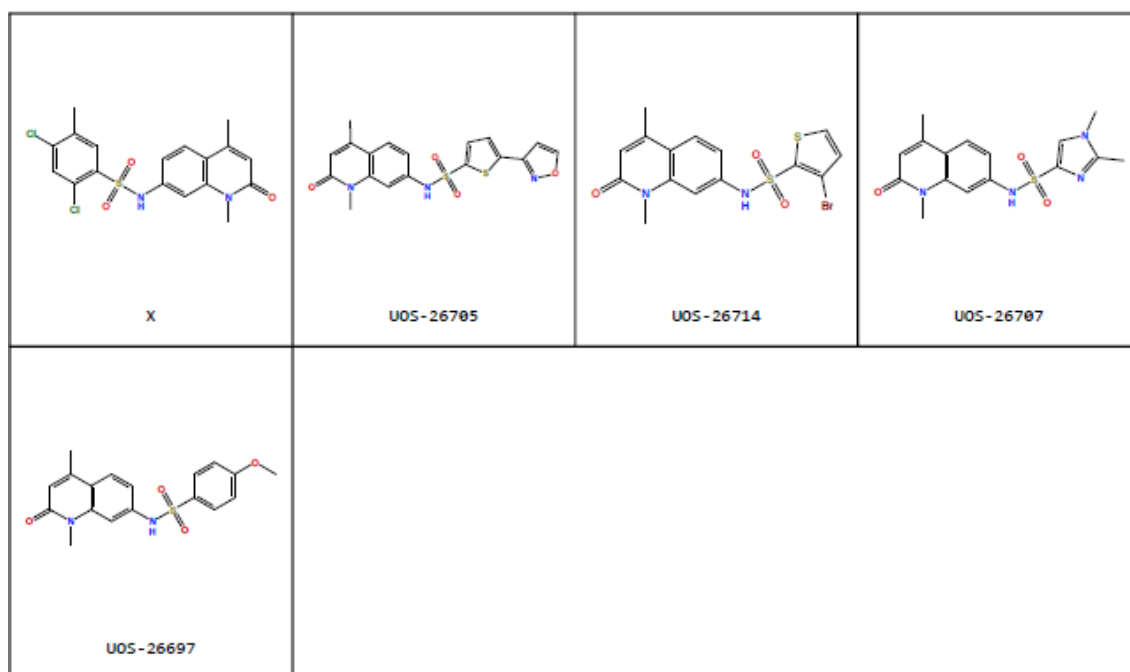
			
UOS-26646	X	UOS-26695	X
			
X	X	UOS-26661	UOS-26662
			
X	UOS-26664	X	UOS-26650
			
UOS-26686	UOS-26657	UOS-26655	UOS-26666
			
X	X	X	X



			
U05-26649	X	X	U05-26656
			
X	X	U05-26690	X
			
U05-26675	X	X	X
			
X	X	X	X
			
U05-26668	X	X	X

**Figure Apx.9. Amides from Commercially Available Acids Selected for Synthesis.** 100 amides selected via docking studies as potential KAT2A and KAT2B inhibitors. Those successfully synthesised labelled with identification codes.

			
X	U05-26702	U05-26713	U05-26703
			
X	U05-26701	X	U05-26706
			
U05-26700	U05-26699	U05-26698	X
			
X	U05-26708	X	U05-26710
			
U05-26711	X	U05-26715	U05-26716



**Figure Apx.10. Sulphonamides from Commercially Available Sulphonyl Chlorides Selected for Synthesis.** 25 sulphonamides selected via docking studies as potential KAT2A and KAT2B inhibitors. Those successfully synthesised labelled with identification codes.

## VII.V. NMR Fragment Screening

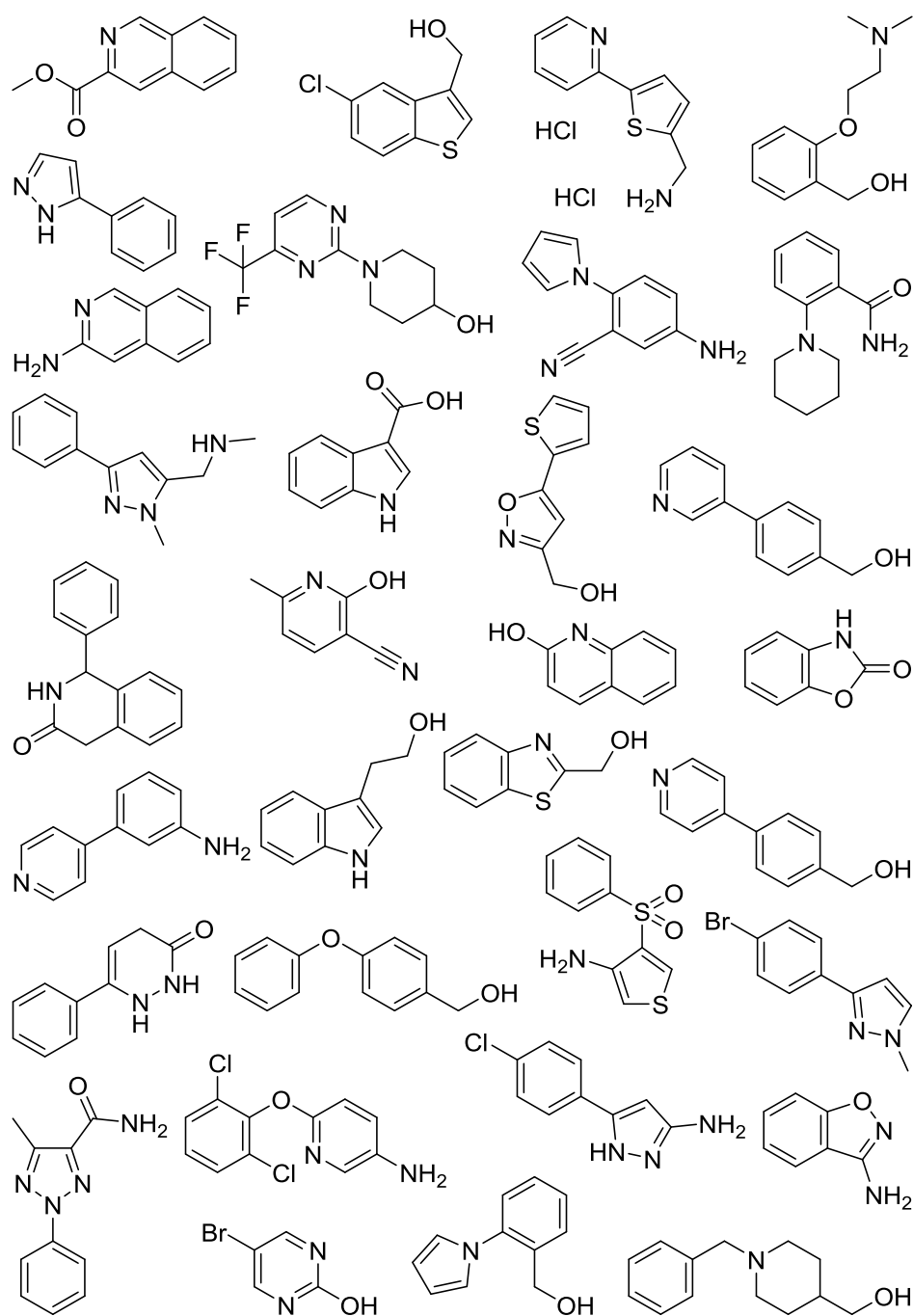


Figure Apx.11. KAT2A BRD Fragment Binders Identified Using NMR Assay.